

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 12:42:18 ; Search time 25.55 Seconds
(without alignments)
5154.826 Million cell updates/sec

Title: PCT-US01-13059-2
Perfect score: 9089
Sequence: 1 MQSIMDSSAVNATEQND.....PRPLMARLHPASKLNKNT 1729

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.68:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3949	43.4	1017	T48452	hypothetical prote
2	2863	31.5	555	T48453	hypothetical prote
3	1559.5	17.2	1207	D84781	hypothetical prote
4	1260	13.9	234	T48454	hypothetical prote
5	709	7.8	917	T05430	hypothetical prote
6	278	3.1	5327	T13564	microtubule-associ
7	246.5	2.7	2897	B48666	cell proliferation
8	246.5	2.7	3256	A48666	cell proliferation
9	240.5	2.6	2938	T30249	cell proliferation
10	233.5	2.6	2218	B84683	hypothetical prote
11	232	2.6	259	D75275	endonuclease III -
12	219.5	2.4	1871	D96796	probable heat shoc
13	217.5	2.4	1490	T20513	hypothetical prote
14	217	2.4	2022	T48818	glucan 1,4-alpha-g
15	211	2.3	1188	T05324	hypothetical prote
16	210.5	2.3	1805	A34736	nestin - rat
17	208.5	2.3	3924	S37431	ankyrin 2, neurona
18	207	2.3	2331	T25410	hypothetical prote
19	206	2.3	3187	JC5837	364K Golgi complex
20	205.5	2.3	2562	T14266	xin protein - chic
21	205	2.3	1983	T00368	KIAA0624 protein -
22	204.5	2.2	2717	A34203	DNA-binding protei
23	197.5	2.2	1145	T18235	transcription acti
24	197	2.2	1366	B86292	hypothetical prote
25	196.5	2.2	3225	T52300	giantin - human
26	196	2.2	3488	T34418	hypothetical prote
27	192.5	2.1	2954	T14156	kinesin-related pr
28	192.5	2.1	3259	A56539	giantin - human
29	192	2.1	1888	T14273	zinc finger protei

30	191.5	2.1	2094	2	S33124	tpr protein - huma
31	191	2.1	1742	2	T49451	kinesin-like prote
32	190.5	2.1	2447	2	T16870	hypothetical prote
33	190.5	2.1	4957	2	T03455	ALR protein - huma
34	190.5	2.1	5262	2	T03454	ALR protein - huma
35	190	2.1	1284	1	WMVZAI	A-type inclusion p
36	190	2.1	2464	1	QRMSP1	microtubule-associ
37	189.5	2.1	236	2	S75373	probable DNA-(apur
38	189.5	2.1	1634	2	T26517	hypothetical prote
39	189.5	2.1	1804	2	T34518	nestin - golden ha
40	189.5	2.1	3498	2	T22330	hypothetical prote
41	189	2.1	2453	2	S60254	nuclear receptor c
42	188.5	2.1	1435	2	A37793	erythrocyte-bindin
43	188.5	2.1	1926	2	JC4842	DNA-binding nuclea
44	188	2.1	1593	2	T22028	hypothetical prote
45	187.5	2.1	264	2	C72770	probable DNA-(apur

ALIGNMENTS

RESULT 1
T48452
hypothetical protein T32M21.160 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48452
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.;
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, March 2000
A:Reference number: 224487
A:Accession: T48452
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1017 <BEV>
A:Cross-references: EMBL:AL162875
A:Experimental source: cultivar Columbia; BAC clone T32M21
C:Genetics:
A:Map position: 5
A:Introns: 167/1; 874/1
A>Note: T32M21.160

Query Match	43.4%	Score	3949;	DB	2;	Length	1017;
Best Local Similarity	99.9%	Pred. No.	2.3e-206;				
Matches	755;	Conservative	1;	Mismatches	0;	Indels	0;
Gaps	0;						
QY	1	MQSIMDSSAVNATEQNDGSRQDVLEFDLNTPOOKPSKRKRKMPKVVVEGPKRKP	60				
DB	235	MQSIMDSSAVNATEQNDGSRQDVLEFDLNTPOOKPSKRKRKMPKVVVEGPKRKP	294				
QY	61	RKPAELPKVVVEGPKRKRKAATQEKVSKETGSAKKNLKEATKKPANYGDMNSKSP	120				
DB	295	RKPAELPKVVVEGPKRKRKAATQEKVSKETGSAKKNLKEATKKPANYGDMNSKSP	354				
QY	121	EYTLSCRKALNFDLENGDARQDSEIVQNSGANSFSEIRDAIGTNGSLDVSQ	180				
DB	355	EYTLSCRKALNFDLENGDARQDSEIVQNSGANSFSEIRDAIGTNGSLDVSQ	414				
QY	181	IDKTNGLGAMNOPLEVSMGNQDKLSTGAKLARDQDPDLTRNQOCQFPVATQTFPME	240				
DB	415	IDKTNGLGAMNOPLEVSMGNQDKLSTGAKLARDQDPDLTRNQOCQFPVATQTFPME	474				
QY	241	NOQAWLQMKNLIGFPFGNQOPRMIRNQOPCLAMGNQOPMYLIGTPRPALVSGNQQLGG	300				
DB	475	NOQAWLQMKNLIGFPFGNQOPRMIRNQOPCLAMGNQOPMYLIGTPRPALVSGNQQLGG	534				
QY	301	POGKRKPIFLNHQTCPLPAGNQLYGSPDTHQVLMSTGGQHGGLLTKNQOPGSLIRGQPC	360				
DB	535	POGKRKPIFLNHQTCPLPAGNQLYGSPDTHQVLMSTGGQHGGLLTKNQOPGSLIRGQPC	594				
QY	361	VPLIDQOPATPKGFTHLNQMVATSMSSPCLRPHSQSQVPTTYLHVESVSRILNGTGTCCQ	420				

Db 595 VPLIDQOPATPKGTHLNMQVATSMSSPGLRPHSQSQVPTTYLHVESVSRILNLTGTCQ 654
 QY 421 RSRAPVDSLOODIHQGNKYILSHEISNGCKKALPONSSLPPIAKLEEARSGSRQY 480
 Db 655 RSRAPVDSLOODIHQGNKYILSHEISNGCKKALPONSSLPPIAKLEEARSGSRQY 714
 QY 481 HRAMGQTEKHDLNLAQAQTAQSQDQVERHNSSTCVYDLAAKTKTKQVVQENLHGMPPPEVI 540
 Db 715 HRAMGQTEKHDLNLAQAQTAQSQDQVERHNSSTCVYDLAAKTKTKQVVQENLHGMPPPEVI 774
 QY 541 EIEDDPTDGARKGNTASISKAGSKGNSPVKKTAKEKCIIVPTPAKKGAGRKKSVPV 600
 Db 775 EIEDDPTDGARKGNTASISKAGSKGNSPVKKTAKEKCIIVPTPAKKGAGRKKSVPV 834
 QY 601 PAHASEIOLMQPTPKPLSPKPKGKRSIQDSGARGSPGELLGODSTAEIILYRMQN 660
 Db 835 PAHASEIOLMQPTPKPLSPKPKGKRSIQDSGARGSPGELLGODSTAEIILYRMQN 894
 QY 661 LYLGDKEREQONAMVLYKGDGALVPYVESKRRKPRKVDIDDETTRIWNLLMGKDEKEG 720
 Db 895 LYLGDKEREQONAMVLYKGDGALVPYVESKRRKPRKVDIDDETTRIWNLLMGKDEKEG 954
 QY 721 DEKDKKKKEWEERRVFRGRADSFARMHLVOGD 756
 Db 955 DEKDKKKKEWEERRVFRGRADSFARMHLVOGE 990

RESULT 2
 T48453
 hypothetical protein T32M21.170 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
 C:Accession: T48453
 R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.; De
 ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
 submitted to the Protein Sequence Database, March 2000
 A:Reference number: Z24487
 A:Accession: T48453
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-555 <BEV>
 A:Cross-references: EMBL:AL162875
 A:Experimental source: cultivar Columbia; BAC clone T32M21
 C:Genetics:
 A:Map position: 5
 A:Introns: 469/3; 496/2; 524/3
 A:Note: T32M21.170

Query Match 31.5%; Score 2863; DB 2; Length 555;
 Best Local Similarity 100.0%; Pred. No. 7.8e-148;
 Matches 549; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 785 SSAPMSLAARPPPKLSSREDERNVRVSVVDDPCCILNLEIPSWOEKVQHPDMEVSG 844
 Db 7 SSAPMSLAARPPPKLSSREDERNVRVSVVDDPCCILNLEIPSWOEKVQHPDMEVSG 66
 QY 845 VDSGSKQLRDCNSGIERFNFLKSLQNEEVLSSQSDSDPAIFOSCGRGVSCSCSKS 904
 Db 67 VDSGSKQLRDCNSGIERFNFLKSLQNEEVLSSQSDSDPAIFOSCGRGVSCSCSKS 126
 QY 905 DAEPPTTRCETKTVSGTSQSVQTSQSPNLSDICLQGNRPHLYEGSGDVQKQETTNVAQK 964
 Db 127 DAEPPTTRCETKTVSGTSQSVQTSQSPNLSDICLQGNRPHLYEGSGDVQKQETTNVAQK 186
 QY 965 KPDEKTMNWKDSVCFQOPRNDTNWQTTPTSSSYEQCATRQPHVLDIEDFGMQGGLGYSW 1024
 Db 187 KPDEKTMNWKDSVCFQOPRNDTNWQTTPTSSSYEQCATRQPHVLDIEDFGMQGGLGYSW 246
 QY 1025 MSISPRVDRVKNKNVPRFRQGGSVPRFTGQIIPSTPHLPGLMGLSGSSSAVOEHQDD 1084
 Db 247 MSISPRVDRVKNKNVPRFRQGGSVPRFTGQIIPSTPHLPGLMGLSGSSSAVOEHQDD 306

QY 1085 TOHQNDQEMKASHLOKTFDLNLSSEELTRQSTKQNTIDGCLPRDRTAEVDVPLSN 1144
 Db 307 TOHQNDQEMKASHLOKTFDLNLSSEELTRQSTKQNTIDGCLPRDRTAEVDVPLSN 366
 QY 1145 NSSLQNTILVESNSNKEQTAVEYKETNATILREMGKTLADGKKPTSQWDSLRKDVEGNEG 1204
 Db 367 NSSLQNTILVESNSNKEQTAVEYKETNATILREMGKTLADGKKPTSQWDSLRKDVEGNEG 426
 QY 1205 ROERKNKNMDSIDYEARRASISEISEAIKERGMNMLAVRIKDFLERIVKDHGSDILEW 1264
 Db 427 ROERKNKNMDSIDYEARRASISEISEAIKERGMNMLAVRIKDFLERIVKDHGSDILEW 486
 QY 1265 LRESPPDKAKDYLLSIRGLGLKSVCEVRLTLHLNLAFPVDTNVGRIAVRMGWVLPLOLPE 1324
 Db 487 LRESPPDKAKDYLLSIRGLGLKSVCEVRLTLHLNLAFPVDTNVGRIAVRMGWVLPLOLPE 546
 QY 1325 SLQJLHLEL 1333
 Db 547 SLQJLHLEL 555

RESULT 3
 D84781
 hypothetical protein At2g36490 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: D84781
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487
 A:Accession: D84781
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1207 <STO>
 A:Cross-references: GB:AE002093; NID:g4581149; PIDN:AAD24633.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g36490
 A:Map position: 2

Query Match 17.2%; Score 1559.5; DB 2; Length 1207;
 Best Local Similarity 28.5%; Pred. No. 1.2e-76;
 Matches 480; Conservative 199; Mismatches 366; Indels 641; Gaps 61;

QY 13 TEATEQNDGSRQDVLEFDLKNTPQOKPSKRRKPKMPKVVVEGPKPKRPAELPKVVE 72
 Db 65 TEEVESLSSVSNVAE-QILKTP-EKPKRKH--PKVRREAKPKREPAPRPRKSVVTD 120
 QY 73 GKPKRPRKAATQEKVK-SKETGSAKKLNKESATKKPANYGDMNSKSPVYTLKSCRKAL 131
 Db 121 QGESKTPKRYVTKKVEVSKD-----QDATPESSAAVETSTRPK---RLCRVL 167
 QY 132 NFDLEN-----PGDARQ-GDSEIVONSGANSFSEIRDAIGTNGSFLDSVSQIDKTN 185
 Db 168 DFEAENGENTNGDIRAEGEMESALQKLDGSGN-QELKDC-----LSAPTPKRKR 219
 QY 186 GLGAMNPLEVSMGNQPDK-----LSTGAKLARDQPDLLTRN-----QQCQFP 229
 Db 220 SQGKRK-----GVQPKNGSNLEEDYDISMAQAARKRQGPCCDMNLSGIQYDEQCDY 271
 QY 230 VATQNTQFPMENQQAQWLQMKNLIGFPFGQOPRMTIRNQOPCLAMGNQOPMYLIGTRP 289
 Db 272 -----QKMHVLYSPNL-----QQGGM--RYDAIC----- 293
 QY 290 ALVSGNOOLGSGPQGNKRPIFLNHQTCPLPAGNQLYGSPTDMHQLVMSTGGQOHLINQK 349
 Db 294 -----SKVFSQOQHNYSVAF--HATCYSSSTQLSANRV-----LTVEER 331
 QY 350 PGSILRQCPVPLIDQOPATP-----KGFHLNQMVATSMSSPGLRPHSHSOV 398

Db 332 EGIFQGRQSELSNLSKIDITDIKKTTGTHARFNLSNMKLV-----EV 376
QY 399 PTYLVHVSRLNGTTTCQSRAPAYDSLQODTHQNKYILSHEISNGCKKALQ 458
Db 377 PE---HL-----TSGYCSKPO-----QNNKILVDTRVT----- 401
QY 459 NSSLPPTIMAKLEEARSGKQYHRAMGQTEKHDLNLAQOIASQDVERINSSTC 518
Db 402 -----VSKKKPTKSEKQ----- 414
QY 519 AKKTKIQVVOENLHGMPEVIEIEDPTDGARKGKNTASISKGASKNGSPVKTA 578
Db 415 ---TK-QKNLLPNLCRFPFSTGLSPD----- 437
QY 579 KCIVPKTPAKKGRAGKRSVPPPAHASEIQLOMPTPKTPLSRSPKPGKGRKS 638
Db 438 -----ELW-----KRRNSI----- 446
QY 639 RGPSELLCQDSIAETIYRMQNLVLDGKEREOONAMVLYK-----GDGALVP 688
Db 447 -----ETISEL-----LRLDINREHSETALVPTTNSQIVLFGGAGALVP 490
QY 689 S-KKKRPKPVDDIDTTRINLLMGKDEKKEKKEWBEERVRFRGRADSEI 747
Db 491 PVKKPRPKVDLDDTDTRVWKLIL-ENINSEGVDSDEQAKAKWEEERVRFRGRADSEI 549
QY 748 ARMLVQGRDRRSPKWSGVSDSVYGVFLTONVSDHLSSAFMSLAARFP-PKLSS 806
Db 550 ARMLVQGRDRRTPKWSGVSDSVYGVFLTONVSDHLSSAFMSLAARFP-PK 608
QY 807 RNVRSVVVEDPEGILNLEISWQ-----EKVQHPDSMEVSGVDSSKQLRDC 859
Db 609 -----AGTSSMPSIQITYLDSSETSSPPDHNHSSVT----- 640
QY 860 GIERFNFLKSIQNLSEELVSSQSDPDPAIFQSCGRVSGSCSKSDAEPPTTRC 917
Db 641 -----LANTQDDEKDVPSNET-----SRSSSEIAISAHESVDKT 676
QY 918 VSGTS--QSVQVQSPNLSDEICLOGNERPHLYEGSDVOK-QETTINVAOKK 974
Db 677 TDSKEVVDSDRKG-----SVEVDKTEKCRVNLNLFPSSEDSALTC 716
QY 975 KDSVCFQGRNDNTNQTTSPSSVEQCATRQPHVLDIEDFGMGEGLGYSWMSIS 1034
Db 717 QHSMVSDAPQNT---ERAGSSS-----EIDLEGE-YRTSFMKL----- 750
QY 1035 KKNKVPFRFFROGGSVPREFTGQIIPSTPHELPGMGLSGSSSAVOEHODDT 1094
Db 751 -----LQGVQVSLSDSNQVSPN-----MSPGDCSEIKGFQ----- 787
QY 1095 KASHLOKTFDLNLSSEE-CLFRQSTSKONITDGCPLPRDRTAEDVVDP 1153
Db 788 KSS-----VDSSEPOCCSQ-----DG-----DVL----- 807
QY 1154 ESNSSNKEOTAVEYKETAETNATILREMGKTLADGKKPTSQWDSLRKQVEG 1213
Db 808 -----SCQKPT---LKEGKKVLKE-----EKAFDCLDRREAAQARAGIR 852
QY 1214 DSIDYEAIRASISEISEAIKRGMMNLAVRI-----KDFLERIVKDHGIDLE 1263
Db 853 DIVDKAIRAADVKEVAETIKSRGNHKLAEIRIQYTLNKKIMQGFGLDR 912
QY 1264 WLRESPPDKADKDYLLSIRGLGLKSECVRLTLHLNLAFLPVDTNVGR 1323
Db 913 WLRDVPPDKAK----- 923
QY 1324 ESLQLHLELYPVLESIQFLPRLCKLDQRTLYELHYOLITFGVKVCTKSRPNC 1383
Db 924 -----YELHYQMITFGVKVCTKSRPNCNACPM 950
QY 1384 RGECHRFASAYASARIALPAPERISLTATIPVPPSPFPVPAIMLEPLPLEK 1443
Db 951 KGECHRFASAFASARIALPSTEGMGTDPKNPLP-----LHLPEPTFOREG 999

QY 1444 PSNREN-----CEPIIEEPASPGOECTEITESIEDIAYNEDDPDIPTIKLIN 1497
Db 1000 VQHSEPAKVKTCCETIIEEPASPEPEAEVSIADIEAEFF-EDPEIPTIRLNMDAFTSN 1058
QY 1498 LREHMERNMELQEGDMSKALVALHPTTTTSIPTPKLNISRLRTEHQVYELPD 1557
Db 1059 LKKIMEHNKELQDGNSSSALVALTAETASLPMPKLNISQRLTEHRVYELPD 1118
QY 1558 DKREPDPSPYLLAIWTPGETANSAQPPQKCGKASGKMGCFDETCSECSN 1617
Db 1119 EKREPDPSYLLAIWTPGETADSIQPSYSTCIFQANGMLCDEETCFSCSNI 1178
QY 1618 RGTLLI 1523
Db 1179 RGTILV 1184
RESULT 4
T48454
hypothetical protein T32M21.180 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48454
R:Bevan, M.; Terry, N.; Ardiles, W.; Buyschaert, C.; Dasseville, R.; De Clerck, R.;
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, March 2000
A:Reference number: 224487
A:Accession: T48454
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-234 <BEV>
A:Cross-references: EMBL:AL162875
A:Experimental source: cultivar Columbia; BAC clone T32M21
C:Genetics:
A:Map position: 5
A:Introns: 49/2; 61/3; 81/1; 127/3; 149/3; 188/1; 201/1
A:Note: T32M21.180

Query Match 13.9%; Score 1260; DB 2; Length 234;
Best Local Similarity 100.0%; Pred. No. 2e-61;
Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1496 MTLREHMERNMELQEGDMSKALVALHPTTTTSIPTPKLNISRLRTEHQVYELPD 1555
Db 1 MTLREHMERNMELQEGDMSKALVALHPTTTTSIPTPKLNISRLRTEHQVYELPD 60
QY 1556 GMDKREPDPSPYLLAIWTPGETANSAQPPQKCGKASGKMGCFDETCSECSN 1615
Db 61 GMDKREPDPSPYLLAIWTPGETANSAQPPQKCGKASGKMGCFDETCSECSN 120
QY 1616 TVRGTLIPCRTRAMRGSFPLNGTYFOVNEFLADHESLKPIDVPRDWIWDLP 1675
Db 121 TVRGTLIPCRTRAMRGSFPLNGTYFOVNEFLADHESLKPIDVPRDWIWDLP 180
QY 1676 SVTISFRGLSTEQIQCFWKGFVGVCFVGFQKTRAPRPLMARLHFFPASKLN 1729
Db 181 SVTISFRGLSTEQIQCFWKGFVGVCFVGFQKTRAPRPLMARLHFFPASKLN 234

RESULT 5
T05430
hypothetical protein F28A23.180 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 24-Nov-1999
C:Accession: T05430
R:Bevan, M.; Weichselgartner, M.; Fartmann, B.; Granderath, K.; Dauner, D.; Herzl, A.
submitted to the Protein Sequence Database, October 1998
A:Reference number: Z15415
A:Accession: T05430
A:Molecule type: DNA
A:Residues: 1-917 <BEV>

A;Cross-references: EMBL:AL021961

A;Experimental source: cultivar Columbia; BAC clone F28A23

C;Genetics:

A;Map position: 4

A;Introns: 41/73; 70/3; 88/1; 118/1; 321/1; 367/1; 474/1; 504/2; 624/3; 655/2; 667/3; 695/3

A;Note: F28A23.180

C;Superfamily: Arabidopsis thaliana hypothetical protein F28A23.180

Query Match 7.88; Score 709; DB 2; Length 917;
Best Local Similarity 21.48; Pred. No. 1.1e-30;
Matches 232; Conservative 108; Mismatches 163; Indels 580; Gaps 22;

```
QY 651 IAEIYRMQNLVLDGEREQEONAMVLYKGDGALVPYESKRRKPRKVDIDDETTRIWNL 710
DB 388 IAKLIKMDRLKINKK-----VTTMIKADKLV-----TAKVLDPEIKEDV 431

QY 711 LMGKDEGEDEKDKKKWEERVRFRGRADSFIAHMLVQDGRFSPWKGSGVSDV 770
DB 432 LM--VNDSPSRSYDDKTEAKWKEREIFQTRIDLFINRMHRLQGNKFKQWKGSGVSDV 489

QY 771 IGVFLTONVSHLSSAFMSLAARFPKLSRSSREDERNVRSVVVEDPEGCILNLEIPSW 830
DB 490 VGVFLTONVSHLSSAFMSLAARFPKLSRSSREDERNVRSVVVEDPEGCILNLEIPSW 830

QY 831 QEKVQHPDMEVSGVDSGSEQLRDCNSGIERPNFLEKSTONLEEVLSQDSDFDAIF 890
DB 531 -----DAKSECI----- 538

QY 891 QSCRGVSGSCSKSDAEPTTRCTKTVSGTSQSVQVGSNLSDEICLOGNERPHLYEGS 950
DB 539 -----ILSDE----- 543

QY 951 GDVQKQETTVAOKKPDLEKTMNKKDSVCFCQPDNDTNWQTTSSSYEQCATRPHVLDI 1010
DB 544 ----- 543

QY 1011 EDFGMOGEGLYSWMISPRVDRVKNKNVPRFRFGSGVPRFTGQIIPSTHELPMG 1070
DB 544 ----- 543

QY 1071 LSGSSAVQEHODTQHNQODEMKNKASHLOKTFDLNLSSECLTROSSYKQNTDGCCLP 1130
DB 544 ---SISKVDHEN-----TAKRKNKGTII-- 565

QY 1131 RDRFAEDVDPVLSNNLSQNLTVESNNSKEQTAVEYKETNATILREMKGTLAGKKPTS 1190
DB 566 -----EDEIVD----- 571

QY 1191 QWDSLRKDVEGNEGRQBRNKNNDSDYEAIRRASISEIAKERGMNNMLAVRIKDFL 1250
DB 572 -WNNLR--MYTKEG--SRPEMHMDSVNNKSVRLSGQNVLETTIKRGQFRILS----- 620

QY 1251 ERIVKHGGIDLEWLRSPDPKADYLLSIRGLGLKSEVCRLLTLHLNAPFVDVTNVGRI 1310
DB 621 ERIL-----VDNTNVGRI 632

QY 1311 AVRMGWVPLPESLOLHLELYPVLESIQKFLWPLRLKLDORTLYELHYQLITFGKVF 1370
DB 633 AVRLGLVPLPLNGVQMH-----QFELYELHYQMITFGKVF 669

QY 1371 CTSRPNACNACPMRGCRHFAFASARLALPAPERSLTSATIPVPPESF-----PPVA 1425
DB 670 CTKTIPNACNCPMKSECKYFASAVVSKVLLESPEEK-----HEPNFMNAHSODVA 722

QY 1426 IPMIELPLPLEKSLASGAPNRCENPEIIEPASPQOECTEITESIEDIAY---NEDPD 1482
DB 723 VDMTSNINLVEECVSGSCSQQAICYKPLVFPFSPRAEIPEST--DIEDVPFNNLYQSYA 780

QY 1483 EPTIKLINIQFGMTLREHM--ERNMELOGDMSKALVALHPTTSTPTP---KLKNISR 1537
DB 781 SVPKIDFDLALKKSVEDALVISGRSSSDEISKALVITPTENACIPKPPKMKMYNR 840
```

RESULT 6

T13564

N;Alternate names: hypothetical protein EG:49E4.1

C;Species: Drosophila melanogaster

C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000

C;Accession: T13564

R;Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.

A;Submitted to the EMBL Data Library, April 1999

A;Description: Sequencing the distal X chromosome of Drosophila melanogaster.

A;Reference number: Z17689

A;Accession: T13564

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-5327 <SPA>

A;Cross-references: EMBL:AL031128; PIDN:CAA20006.1

C;Genetics:

A;Cross-references: FlyBase:FBgn0025392

A;Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1

A;Note: EG:49E4.1

C;Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match 3.18; Score 278; DB 2; Length 5327;

Best Local Similarity 19.7%; Pred. No. 3.3e-06;

Matches 277; Conservative 198; Mismatches 583; Indels 348; Gaps 59;

QY 2 QSTMDSAVNATATEQNDGSRDVLFDLNTPOOKPSKRKFMKPVVVEG-----K 55

DB 1986 ESKD-----EAKSKEESRRE-----SVAEKSPLPSKEASRPASVAESIKDAEK 2031

QY 56 PKRKRPAELPKVYVEGPKRKP-----RKAATQEKYKSKETGSAKKNNKESATKK 108

DB 2032 SKEESRESVAEKSPLPSKEASRPASVAESIKDAEKSKEESRESAAEKSPLPSKEASR 2091

QY 109 PANVG-----DMSNKSPEVTILKSKRKALNFDLENPGDAR--QGDSE--SEIVQNSGANSF 160

DB 2092 PASVAESVKDEADKSKEESRRE-----SMAESGKAQSIKQDQSPLEKVSRRPESVAESV 2144

QY 161 SEIRDAITGGTNGSFLDSVSOIDKTNGLGAMNPOKLSGTGAKLARDOQPDLL 220

DB 2145 KD--DPVKSEPRRESVAGSVTADSDQSPLESGASRPESVVDVDEAEKQESRR 2202

QY 221 TRNQOCQFPVATQNTQPMENOQAWLQMKNLQIGFPGNQOQPRMTIRNQOQCLAMGNOQP 280

DB 2203 ESKTESVIPPKAQDKDKSPKEVLQ-----PVSMTETIREDADQ-----DQ 2240

QY 281 MYLIGTPRALVSGNQO-----LGGPQGNKRPIFLNHOTCLPAGNOLYGSTDMHQLV 333

DB 2241 M-----KPSQAESRRRESIAESIKASPRDEKSPLEASKEASRPESVAESIKYDLDKPKII 2294

QY 334 MSTGGQOH-----GLLTKN-----QPGSLRGQOPCVPLID--QOAPATPK-----GF 374

DB 2295 KDKKSTEHSSRESLEDKSAVTSKSVSRPLSVASDHEAAVAIEDDAKSSISPKDKSRPGF 2354

QY 375 THLNQMVATSMSSPGLRPHRSQSQVPTTYLHVESVSR-----TLNGTTGTCQRSRAPAYD 428

Db	1829	AQPLEDLAGLKFQPPVCTDPTTHEKTKTIACKSPQDPVGPPTTIFKPOSKRSLRKAD	1888
Qy	700	IDDETRIRIWNLMKGDEKE-----GDEKDKKEKWEERRRVERGRADSFARMHL--	752
Db	1889	VEESLALRKRTPSVGKAMDTPKPAGGDEKMK-----AFMGTP---VQKLDLPG	1935
Qy	753	-VOGDRRFSPW-----KGSVVDVSVIG---VFLTMQVSDHLLSSAFMSLAARFP-----	796
Db	1936	NLPGSKR--WPQTPKEAQALEDLAGFKELFQTPGTDPKTDTTEKTKTIACKSPQDPVD	1992
Qy	797	PKLSSREDERNVRSVVVEDPGCCILNLNEPQWEKVOHQHPSDMEVSGVDSGKEQLRDC	856
Db	1993	TPASTKQRPKNRLKRAVDE--EFLALRKRTPSAGKAMDTP-----KPAYSD	2038
Qy	857	SNGIERFNFLEKSIQNL-----EVLSSQDSFDPAPFQSCG-----	894
Db	2039	KNIN-----TFVETPVQKLDLLGNLPGSKRQPOTPKEAKALEDLVGF-KELFQTPGHTEE	2093
Qy	895	-----RVGSCSKSDAE-FPTTRCE-----TKTVSGTSSQSVQTG	928
Db	2094	SMTDDKITEVSSQPOPESEFKTSRSQKQRLKPLVKVDMKEEPLAVSKLRTSGETTQTH	2153
Qy	929	SPNLSDBIQCLOG-NERP-HLYEGSDV-OQETTINVAQKPDPLEKTMNKWDSVCFQOPR	984
Db	2154	TEPTGDSKSIAKAPESPKQIILDPAASVTSGRQLRTRKEKARALEDLVDFKE--LFSAPG	2211
Qy	985	NDTNWQVTPSSYEOCATRPHVLDTEDFGMGEGLGYSWMSISPRVDRVKNKNVPRRF	1044
Db	2212	HTESMTIDKNKIPCKSPPELDT-----ATSKRCPKTR---PRKEV	2253
Qy	1045	ROGGSVPREFTGOIPST-PHELPGMGLSGSSAVQE-----HODDTHOQNQDEMNK	1095
Db	2254	KEELSAVERLTQTSQGSTHTRPEASGDGKIVLKORAKKPNPVEEBSRRRPAPKEK	2313
Qy	1096	ASHLOKT--FLDLLNS-----ECLTRQSTQNITDGLCPDRDRAEDVVDPLSNSSLQN	1150
Db	2314	AQPLEDLAQFTELSETSGHTQESLTAGKATK-----IPCESPPLVVDVDTASTKRHLR	2366
Qy	1151	ILVESNSSNKEOTAVBYKETNATI-----LREMKGFLADGKKPTSQOWSLRKDVDGNE	1203
Db	2367	TRVQKVQVVEEPSAVKFTQSETTDDAKEPAGEDKGKA-LKESAKQTPAPAASVTGSR	2425
Qy	1204	GROERNKNMDSIDYEAIRRASISEISAIRGMNNMLAVRIKDFLERIVKDHGGIDLE	1263
Db	2426	RRPRAPESAQAIE---DLAGFKDPAAGHTEESMTDDKTKIPCKSSPELDTATSS--	2479
Qy	1264	WLRESPPDKA-----KDYLLSTRGLGKSVCEVCRLLTLUHLNLAFPVDNVRGIARVMGWVP	1318
Db	2480	--KRRPRTRAQVKEELLAVGKLTQTSGE---TTHTDKEPVGECKGTAKF-----	2527
Qy	1319	LQPLPSLQHLLELYPVLESIQKFLWRLCKLDORTLYEL-----HVQLITFG	1367
Db	2528	-QPAKRN-----VDADVITGSRRQ---PRAPKEAQLPDLASFOELQSTPGHTEELANG	2578
Qy	1368	KVFCTKSRPNCACPMRGCEHFASAYASARLALPAPER---SLTSATIPVPPEPPVP	1424
Db	2579	AADSFTSAP--KQTPDSGK-----PLKISRRLRAPKPEVPVGVSTROPVRSQKSNT	2630
Qy	1425	AIPWIELPLPLEK---SLASGAPSRENCEP-----IIBE-PASPGQCTEITESDIEDAY	1476
Db	2631	SLP---PLPFRGGGKDGSVTGTRLRCMPAPEIIVBELPASKQR-----VAPRA	2678
Qy	1477	YNEDPDEITIKLNTIEQGMTLREHWERNMELQECDSKALVALHPTTTSIPTPK-----	1531
Db	2679	RKSSSEPVVIMKRSR-----TSAKRIEPAEELNSNDM-KTNKEGHKLQDVSPEKNGLSLR	2733
Qy	1532	LKNISLRTEHWOVEL-PDSHRLLDGMDKREPDDPSPYLLAIWPPGETA-----	1579
Db	2734	SRQDKTEAQIITEVFLAERIENREKKPKMTSPD-MDIQNPDDGAKRPIPROKVTE	2792
Qy	1580	-----NSAOPP-EQKGGKASGKMGCFD-----ETCSNCSLREANSQTVR	1618

Db 2793 NKRCLSRQARNESSQPVKVAESGGQKSAKVLMQMKGKSGANGSDSMCLSRKTSPAA 2852

Qy 1619 GTL 1621
||

Db 2853 STL 2855

RESULT 8
A48666
cell proliferation antigen Ki-67, long form - human
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 08-Oct-1999
C:Accession: A48666
R;Schlueter, C.; Duchrow, M.; Wohlenberg, C.; Becker, M.H.G.; Key, G.; Flad, H.D.; Ge
J. Cell Biol. 123, 513-522, 1993
A:Title: The cell proliferation-associated antigen of antibody Ki-67: a very large, u
ins.
A:Reference number: A48666; MUID:94043435
A:Accession: A48666
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-3256 <SCH>
A:Cross-references: EMBL:X65550; NID:g415818; PIDN:CRA46519.1; PID:g415819
C:Superfamily: kinase interaction domain homology
C:Keywords: alternative splicing; cell cycle control; nucleus; tandem repeat
F:29-91/Domain: Kinase interaction domain homology <KI>

Query Match 2.7% Score 246.5; DB 2; Length 3256;
Best Local Similarity 18.8%; Pred. No. 8.3e-05;
Matches 362; Conservative 266; Mismatches 768; Indels 527; Gaps 94;

Qy 22 SRQVLEFDLNKTPQQKPSKRKRMPKVVVEGK----PKRKPR--KPAELPKVVVEGKP 75
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1496 SQPDPVDTPSSKKPQSKSLR-----KVDVEEFFALRKRTPSAGKAMHTPPAVSGEK 1549

Qy 76 KRKPRKAATQEKYKSTGTGSACKKNL---KESA-----TKRPANVGDMSNK 118
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1550 NIYAFMGTPVQKLDTENLTGSKRRRLQTPKEKAQALEDLAGFKELFOTRGHTEESMTNDK 1609

Qy 119 SPEVTLSCKRALNFEDLENPCDARGDSEFIYONSSGANSFSIRDAIG---GTNG--- 172
: | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1610 TAKVACKSSPDLD---KNPASSKRR-----LKTSLGKGVGRELLAVGKLQTGSTETT 1660

Qy 173 -----SFDDYSQI-DKTNGLGAMNOPLEVSMGNQ--PDKLSTGAKLARDQ 215
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1661 HTHTPTPGDGKSMKAPFESPQILDSASITGSKRQLRTPGKSEVPEDLAGFIELF-Q 1718

Qy 216 QPDLLTRNQCOFFPATONTQFPMENOO-----AWLMQNOLIGFPFGNQ-QPRWtir--- 267
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1719 TP-----SHTKESMTEKTTKVSYRASQPDLDVTDPTSSKQPQRSLRAD 1763

Qy 268 NQOPLANGNOQM--YLIGTPRALYSGNOOLGGPOGNKRPIFLNHOTCLPAGNOLIYS 325
: | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1764 TEEEFIAFRKQTPSAGKAMHTPKPAV-----GEEKDI-----NTFLGT 1801

Qy 326 PTDMHQLVMSTGGQOHLLTNQOPGSL--IRG-----QOPCV--PLIDQQPATPKGFTH 376
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1802 PVQLDQPNLPGNSRRLRQTRKEKAQALELTGPREFLOTPCTDNPTADEK-----TT 1854

Qy 377 LNOMVATSMSPGLRHRSQSQPTYTLHVESVS-----RILNGTTCORSRAPAYDSL 430
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1855 KKLICKSQSPDADPTNTKQRPKRSLLKADVEEFLAEFRKLTPSAGKAMHTPKAAVGE- 1913

Qy 431 QODTHQ--GNKYILSHISINGNGCKKALPONSSLPTIMAKLEEARGSKQY----HRAM 484
: | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1914 EKDINTVGTGPVEKLDLLGNLPGRKR-POT--PKEKALEDLAGFKELFTPGHTEE 1969

Qy 485 GOTEKHDNLNAQQIAQSODVERHNSSFCVEYLDAAKTKIQKVVENLHCMPEVEIED 544
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 1970 SMTDDKITEVCSQSPDPVKTPTSS-----KORUKISLGKV-----GVKEEVLPV-- 2015

Qy 545 DPTDGARKGNKTASISKGA-----SKGNSSPVKRTAKEKCIV-----PKTPP 587

QY 364 IDQAPATPKGTHLNMVATSMSPG-----LRPHSQVPTTYLH--VESVRILNLTG 417
Db 1562 LFOTPSGKDPVTVDKTKIALQSPQGHIIINPASKRQSNMRLKDMREFSILEKQTQS 1621
QY 418 TCQRAPAPVDSLOODIHQNKYILSHE-----ISNGNCKK--ALPONSSLPPTIMAK 469
Db 1622 RGRDAGTPA--PMQEE--NGTTAIMEPTKOKLDFIGNSTGHKRRPTPKRAQP----- 1671
QY 470 LEEARGSKROYHRAMGOTEKHDLNLAQOI--AOSQ--DVERHNSSTVEYLDAAKTKIKQ 526
Db 1672 LEDLDGQELFOTPAGASDPVSVESAKISLASSQAEPVTPAST-----KRSKTKGLSK 1726
QY 527 VVQENLHMPPEVIEIEDPTDGARKGNKTASTISKGASKNSPVKKTAEKCIIVKTP 586
Db 1727 V-----DVRQEPSTLGRMK-----SLGRAPGTPAPVQOE--ENDSTAFMETP 1766
QY 587 AKK-----GRGRKKSVPVPAHASEIOLWQPTPKTPLSRSPKPKGKRSIQD-----SGK 637
Db 1767 KOKLDFTGNSGKRR-----POTPKIRAQP-----LEDLDGQEL 1802
QY 638 ARGPSGELLQODSIAEIIYRMQNLGLDKEREQEBONAMVLYKGDGALVPYVESKK--RKPR 695
Db 1803 FOTPAG--ANDSVT-----VEESVKMSLESSQAEPVTPASTKRLSKTGL 1845
QY 696 PKVIDDITRIWNLMLGKDEKDEKDKKKEWEEERRVFRGRADSFAR----- 749
Db 1846 SKVDVREDPSILEKTKTSPTPAPVQENDCTAFMETPKOKLDFTGNSGKRRPTPKI 1905
QY 750 ---MHLVQGDREPSKGSVDSVIGVILTONVSHLSSAFNSLAARPPKLSRSSRED 805
Db 1906 RAQPLELDGQELFOTPAGASDSV-----TVEESAKMSLESSQAEPVTPASTKRLS 1958
QY 806 ERNVRSV--VVEDPEG--LILNLEIPSWQEKVOHPSD-----MEVSGVDSGSK 850
Db 1959 KTGSLKVDVREDPTSLGKTKSPGRACGTPAPVQENDCTAFMETPKOKLDFAENSSGSK 2018
QY 851 EQLRDCNSGIERENFLEKSIQNLLEEVLSSQDSFDPFAIFQSGRGVSCSCSDAEFFT 910
Db 2019 RRSRTSKN---RSQPLE--DLDGQELFOTPAGASNPVSVESAKI---SLESSQAEPVR 2070
QY 911 TRCTKTVSGT---SQSVQVGSPLNLSDEICL-----QGNF----- 942
Db 2071 TRASTKRLSKTGLNKMVDVREHSPKSLSSCASQKVMQTLTLGEDHGRETKDGKVLAKL 2130
QY 943 RPHLYEGSGDVQKQETTNVAKPDL-----EKTWNKDSVCFGQPRNDTNMOTTPSS 995
Db 2131 EPALIVYTRGRQORSCKKRSQSPEDLSGVQEVFQTSCHKNSDVTV-----DNLAKLPSS 2184
QY 996 SYEQCATRQPHVLVDIEDFGMGEGGLGYSNMSISPRVDRVKNKNVPRFRQGGSVPREFT 1055
Db 2185 S-----PPLETDTSVTSRRQARTGL-----RKVHKNE-----LSGGIMHPQIS 2224
QY 1056 GQIITPSTPHELPGMG--LSGSSSAVQEHODDTQHNQODENMKASHLOKTFDL-- 1106
Db 2225 GEIV--DLPREPEGEKVIKTKOSVKRKLDETVNVPKSRORITRAEKTLEDLPFGQELC 2283
QY 1107 -----LNSSECLTRQSSSTQNTDGLCPDRDAEDV-- 1138
Db 2284 QAPSLVMSDVIVEKTPKMPDKSPPEVDTTSETQARRLRRLVLVTEEPORKTTRVVRQT 2343
QY 1139 -----VDPLSNSSLONL-----VESNSNKEQTA- 1164
Db 2344 RNTQKEPISDNQMEEFKSSVQKQDPSVSLTGRNRQNPRTVKEQTQLELTQFEETAK 2403
QY 1165 -----VEVKETNATILREMGTLA--DGKK--PTSOWDSLKQDVEGNROERNKNNM 1213
Db 2404 RISKSQPBEKETLAGKROLRLQLINDNGKKEPTAQ-----RKQ--PSRETRNTLKEPVG 2458
QY 1214 DSDIYEAIRBASISEAIEKRGKNNMLAVRIKDFELIRIVKHG--CIDLEWLR----- 1266
Db 2459 DSNIVEEVKSTQKI-----DPVASVPVSKRRPRVPEKAQALELAGLGPQITL 2509
QY 1267 -----ESPDK-----AKDYLLSIRGLGLKSVE 1289

Db 2510 GHTDESASDRGPTQPCNSLQPEQVDSFQSPRRPRTRRGKVEADEEPSAVR-----KTVS 2565
QY 1290 CVRLTLTLNLAFVDTNVG-----RIAVRMGW--VPLQPLPE 1324
Db 2566 TSR--OTMRSRKRVFEIGNNGTQVSKASIKQITDVAKVTGSRRLRTHKGWSTILLKLLGD 2624
QY 1325 SLQI-----HLLLELYPVLESIQK-----FLWP--RLCKLDORTLYELHYOLITFGKVFC 1371
Db 2625 SKEITQISDSEKLAHDTSLKSTQOQKPDVSKPLRTRCRRVLRSKKEVPKEVLVDTRDHA 2684
QY 1372 T-KSRPNACNACPMRGECRHFASAYASARLAL-----PAPEERSLSATIPVPPE 1420
Db 2685 TLOSKNPLSPKRSKARDGSIVTRALRSAPKQEADEKVPKRAASRKYVSE- 2743
QY 1421 FPPVAIPMIELPLPLEKSLAS-----CAPSNRENCEPIIEEPASPGOECTEITESDIEDA 1475
Db 2744 --PVKMKHLKI---VSNKLESVEEQVSTVMKTEMEAKRENPTVTDQ-----NS 2787
QY 1476 YNEDPDDEIPTIKLINIEQFGMTLREHMERNMELQEGDMSKALVALHPTTTSIPTPKL--- 1532
Db 2788 RYRK-----KTNVKQ-----PRKPEDAS 2805
QY 1533 -----KNISRLRTEHQVYELPDHRLDGMKREDDPSYLLAIWTGCTANSQPP 1585
Db 2806 AENVGIKKNEKTKMTASOETEL-----QNPDD-----GAKKTSR-- 2840
QY 1586 EQCGGRASCKMCFDETCSECSILREANSQTVRGCTLLI---PCRTAMRGSP 1634
Db 2841 -----GOVSK-----LRS-----RGTEMPOPCAEKTSKP 2871
RESULT 10
B84683
hypothetical protein At2g28300 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: B84683
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: B84683
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2218 <STO>
A:Cross-references: GB:AE002093; NID:g4803953; PIDN:AAD29825.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g28300
A:Map position: 2

Query Match 2.6%; Score 233.5; DB 2; Length 2218;
Best Local Similarity 19.1%; Pred. No. 0.00024;
Matches 345; Conservative 252; Mismatches 644; Indels 563; Gaps 88;

QY 8 SAYNATEATEQNDGSRQDVLEF---DLNKTPOQKPSKRRKFMFKVVEGKP--KRKPRK 62
Db 136 SAYSVRTQATNAISSAATGLDFVSSDKRLEAAASHPTSSSLALTSPD---LSGPPGFQSLPAS 193
QY 63 PAELPKVVVEGPKRKRKAATQEKVKSKETGS-----AKKNLKE 103
Db 194 PATP-IRGRGRSRRGAGRRVGVHLSNSSITQRTETATSLASDAEATKALPR 252
QY 104 SATK-----KPAVGMDSNKSPEVTLKSCRKALNFD--LENPGDAQOG--DSES-----EI 150
Db 253 SASEIVSRVPKANEGSTSNPDQVSPVHSATTALRSKADKDLDPGDFSGSHVQTLNV 312
QY 151 VQNSSGANSFS--EIRAIGQ---TNGSFLDVSQIDKTNGLGAMNPOLEVSMGNQPKL 205
Db 313 LENSSEKFAVAKRRLPIQGGGPVQONVSSVCDGSKSPSEGRITYALQ--GVTTAP--- 368

Db 198 -----DPPYLHELHNSHGRQVCTWTRPKGKILRERC-----DAYAL 238
QY 1397 ARLALPAPEER 1407
Db 239 YGDKVPSPSEK 249

RESULT 12
D96796
probable heat shock protein, 53413-59028 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: D96796
R:Thellogis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: D96796
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1871 <STO>
A:Cross-references: GB:AE005173; NID:g6143906; PIDN:AAF04452.1; GSPDB:GN00141
C:Genetics:
A:Gene: F28016.15
A:Map position: 1

Query Match 2.4%; Score 219.5; DB 2; Length 1871;
Best Local Similarity 17.2%; Pred. No. 0.0011;
Matches 251; Conservative 222; Mismatches 510; Indels 477; Gaps 61;

QY 86 EYKSKETGSAKKNLKESATKPKVANGDMNSPKSYTLKSCRKALNF-DLENPGRAROG 144
Db 103 ERTEEEKGLADSNKESVDSLRLP-----PDTEGRECHEQTRHEQENNKQLVQA 152
QY 145 DSESEIVONSSGANSFEIRDAIGTNGSFLDSYSQIDKTNGLGMNQPLEVSMGN---- 200
Db 153 ESDD---SDDFGSRAFEIEE-----QESDVLDRSTSGAMEKETDQVGLR 199
QY 201 -----QPKLSTGAKLARDQPDLLTRNQOCQFPVATQNTQFPMENQOAWLQKQLIGF 255
Db 200 VQIEEPERHNEESKISE-----MVDGETSGHEKKKVVKM 234
QY 256 PFGNQOQPRMTRNQOCPCLAMGNQOQPMYLICTPRPAL-----VSGNQQLGGPQGNKRPIFLN 311
Db 235 DKNRDVKEVDG-----AMGEF-----RPNIDRTQVVGDEDAETAEKNDDEP--- 278
QY 312 HQTCPLAGNOLYGSPTDMHOLVMTSGQGLHLIKNQPSGLINGQOQCPVLIDQOQATP 371
Db 279 -----ESDLEADEVK-----INEGNTK---VRHSEDRNLKLOEK-----EQHSKEQ 322
QY 372 KGFT---HLNQVATSMSSPCLRPHSQSOVPTTYLHVESYSRLINGTTTCQSRAPAYD 428
Db 323 KGSKEENMKELV-----EETP-----EAETTRNDILGPGQIEVPEVD 363
QY 429 SLOQDIHQG---NKVILSHSISNGCKKALPQNSSLPTPTMA----- 468
Db 364 TLGKTSDEGKQNIYKELKNGDATEIDAKMGVAFASNIADTGMNSEDFFESDKLESAD 423
QY 469 ---KLEPARGSKQYHRAMQTEKHDNLNLAQITA---QSQDVERHNSSTCVLEYL-----D 517
Db 424 EVDKMWKDKROENDKVGQAQSEDISLTKIQLGEIQEQFQCKRHKDKOENIKELREGQASE 483
QY 518 AAK-----KTKLOKVQVENLHGMP-----PEVIE-----IEDDPT----- 547

Db 484 AEKNIKNDILKVPQKRSEGHKHKIQKTFQEETNKQPEGYNEKIMETGKKINEDGTRKQVEM 543
QY 548 -----DGARKGKNTASISKGASKGNSSPVKTAETAEKCIKVPKTPAKKGRAGRKSVPP 600
Db 544 IROQELDEPARSEKENRSRELKSVKTNDEKK-----EKE-----IAGTERKEKESDRPKILRE 597
QY 601 PAHASEIQLWQTPPKTPLSRSPKPKGKRSIQDSGKARGPSGELLQCDSIAEIIYRMQ- 659
Db 598 QEVADEV-----AEDTKFSIYGEVEKEEELAGKEFEFGSDDDIARIVRTEQ 645
QY 660 ---NLYLGDKEREQEQNAMVLYK-----GDGALVPYSEK-----KRPRPKVDIDDE 703
Db 646 LDSNAMQGEQEKMIQELVLEEKVCDGKGIIVAAETKAENKSKRVQETEEQKLDKEDT 705
QY 704 TTRWNLLMGKDEKDEKDEKDKKKKXWEEERV-FRGRADSFIAARMHLVQGRDRSPW 762
Db 706 CGHFQKLIIGEISDHGEVEDEKGRKTAERKIKRDRAREAEAEIKDKDLGVSGRYI--- 762
QY 763 KGSVVDSDVI---GVFLTONVSDHLLSSSAFMSLAARFPKLLSSSRREDERNRSVVVEDEP 819
Db 763 KGTIKELVENRGIYRNEH-----EKKKDDAN-----RPEK 794
QY 820 CILNLEIPSQEKKVQHPSDMEVSGVDSGK-----EQLRDC-SNSGIE-----RFNLE 868
Db 795 ITGTIK-----QELVSLNSQLROENVEDGKTQBELVEEKIKDCDEERGSEESKIKTDVV 849
QY 869 KSIONLEEVN-----SSQDSFQPAIFQSGRVGSCSCSK 903
Db 850 RKVOGIEEELYKPKRPHGKTKITELVEETGDIYKQEKETAESDIEAECG-----SLRK 904
QY 904 SD-----AEFTTR-----CETKTVSGTSQSQVQTGSPN 931
Db 905 VDGIEEHHEPETHKERDNNRVTGAKPSQGEKEKEKIVESTMITENDNSIDVQETK 964
QY 932 LSDEICLOGNERPH-----LYEGSGDVQKQ-----ETTNAOKK----- 965
Db 965 KERPGRLSHDKRYKIOELLMEAGHNDKREEQNENYTAEVELETSRVSKKVOEGKMD 1024
QY 966 -----PDLEKTMW-----KDSVCFQOPRND-----TNWQ 990
Db 1025 DNSGKFHEFEERKSYEDWTHEKREKRVLYEEETYPKDKHTGEGDNDHKEEQKENV 1084
QY 991 TTPSSSYCATROPHVLDIEDFGMOGEGLYGWSMTSPRVDRV-----KNKNVPRRFRQ 1046
Db 1085 AKAELNTEEDSFKAKEIEKHODGELKRSVMVQAKRQETEEKDTRAMEKNETVERRKQTK 1144
QY 1047 GGSVPREFTGQIIPSTPHELPGWGLSGSSSAVQV-----HODDTQHNOQDEMKNASH 1098
Db 1145 DGSU-----SKLREGEDPELGGHRRGEEDRIEELVETESDHKEKVKKDEDIYLSQD 1199
QY 1099 LQKTFDLNLSSEECITRQ---SSTKQNTDGLCPDRDRTAEDVDVPLSNSSLQNLVESN 1156
Db 1200 TGR---VDLGERERRSKQTHKSVEDEIGD---QEDEDAEAAVVSNN-----ENG 1246
QY 1157 SSNKQTAVE-----YKTN-----ATILREMGTLADGKKKPTSQWSDSLR 1196
Db 1247 SSRKQVTIEESEKHKQKQKIPETSNPEVNEEDEERVVEKETKEVEAHVQLEGGTENC- 1305
QY 1197 KDVEGNEGROERNKNNMDSIDYEAIRASISEAIEKERGMNNMLAVRIKDFLIRVKD 1256
Db 1306 KDDGEGRREREKQGMTA-----ENMLRQREK-----TKS 1336
QY 1257 HGGI---DLEWLRESPPDKAK 1274
Db 1337 DDGIVRKIQETKEEPEDEK 1356

RESULT 13
T20513
hypothetical protein F02E9.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T20513

R.Gray, I.
submitted to the EMBL Data Library, November 1996

A:Reference number: Z19285

A:Accession: T20513

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1490 <WIL>

A:Cross-references: EMBL:Z81494; PTDN:CA04052.1; GSPDB:GN00019; CESP:F02E9.4

A:Experimental source: clone F02E9

C:Genetics:

A:Gene: CESP:F02E9.4

A:Map position: 1

A:Introns: 225/3; 321/2; 373/3; 474/3; 619/3; 664/3; 828/3; 868/3; 902/2; 965/2;

Query Match

Best Local Similarity 2.4%; Score 217.5; DB 2; Length 1490;

Matches 315; Conservative 221; Mismatches 555; Indels 547; Gaps 86;

```
QY 281 MYLIGTPRALVSGNOQLGGPGNCRP-----IFL-----NHQTCLPAGNQLYG 324
  || || || || || || || || || || || || || || || || || || || || ||
Db 1 MY---NPPGGGGGNN--GGDSQQQQTNNATFLFLQMIQSQHQHQHON--QQQQQLEL 53
  || || || || || || || || || || || || || || || || || || || || ||
QY 325 SPTDMHOLVMTGGQOHGLLKNQPGSLIRGQPCVPLIDQOPATPKGFTHLNQWATS 384
  || || || || || || || || || || || || || || || || || || || || ||
Db 54 QTRDERILIEQRMQH-----QQQONQLQG-----LNQFFNPLGLFQVQAQAA 101
  || || || || || || || || || || || || || || || || || || || || ||
QY 385 MSSPLGRPHSQ--SOYPTTVLHVESVSRIILGTGTCORSRAPAYDSLOQDIHQGNKIYLS 443
  || || || || || || || || || || || || || || || || || || || || ||
Db 102 QAA--FAQNAQSPIP--FHIGS-----PLQPSHPAASALQO-----QYL-- 138
  || || || || || || || || || || || || || || || || || || || || ||
QY 444 HEISNGCKKALPONSLPTPIMAKLEARGSKR-----QYHRAMGQTEKHDLNIA 495
  || || || || || || || || || || || || || || || || || || || || ||
Db 139 -----LPSHPAITFARNSEARNIEQFIAQEEAANVRANSQQQS---PLI 183
  || || || || || || || || || || || || || || || || || || || || ||
QY 496 QOIAQODVERHN---SSFCVYLDAAKTKIQVQENLHGMPEVIEIDPTDGARK 552
  || || || || || || || || || || || || || || || || || || || || ||
Db 184 RPIQQOALNIONLSTQAOQILAHHRQVPVQVQVQHQHIIPTPLALPI-----AQO 236
  || || || || || || || || || || || || || || || || || || || || ||
QY 553 GKNTASISKASKGSSPKYKTAKEKCIQVPTPA-----KKGRAGRKSKSV 598
  || || || || || || || || || || || || || || || || || || || || ||
Db 237 G-----PISNEVSPVPVVPATSGCPQPREPQQGGRRQNRPGRRKKP 280
  || || || || || || || || || || || || || || || || || || || || ||
QY 599 PPPAHASEIQLWOPPTKPTLSRSPKPKGRKSIQDSKARGPSGELLCQDSIAEIIYRM 658
  || || || || || || || || || || || || || || || || || || || || ||
Db 281 EGPPRVDEALAYLRVIKSTFSFSDVPVYHRLFKDKFRAQRIETPDVI--EQVAELLYDS 338
  || || || || || || || || || || || || || || || || || || || || ||
QY 659 QNLXLG-----DKE-----REQQNAMVLYKGDGALVPYES 689
  || || || || || || || || || || || || || || || || || || || || ||
Db 339 PELVGFNFPLTGYRITLTPDKKVVFSQPQMPRVLLSPDERRARAIEAQAQVGAIEL 398
  || || || || || || || || || || || || || || || || || || || || ||
QY 690 KRRKPRPKVIDDETTRIWNLLMGKDEK-----EGDEE---KDKKKEKWEEERRV 738
  || || || || || || || || || || || || || || || || || || || || ||
Db 399 GSQEGISK---DEDTIEMDMKSKEKDDYDGDIDDEDESGIEDKNNEEMEDNHL 454
  || || || || || || || || || || || || || || || || || || || || ||
QY 739 FNGRADSFARMLVQGRRRFPWKGVSVDVIGVFLTONVSDH-----LSSAFMSIA 792
  || || || || || || || || || || || || || || || || || || || || ||
Db 455 IE-----EIIICDDRK---KDDCEDSQQEMSELSAHTLNIITELKKKSF--- 497
  || || || || || || || || || || || || || || || || || || || || ||
QY 793 ARFPKPL-----SSREDERNRVSVVVEDEGCI-----LNLNIP---SWQKV 834
  || || || || || || || || || || || || || || || || || || || || ||
Db 498 AR-PTKLVDFMTFIDFMSDQYKDKMLRKDDDEDEIEENEKIEVDVDPGSPNAPQEI 556
  || || || || || || || || || || || || || || || || || || || || ||
QY 835 QHPSDMEVSGVDSKQELRD--CSNSGIERFNFLEKSIQNL---EEVLSSQDSFDPALF 890
  || || || || || || || || || || || || || || || || || || || || ||
Db 557 KPDDIEKK--DSSKNLQIESSCS-----DYLVSMLANCCIGPDLAAITDFLPI- 606
  || || || || || || || || || || || || || || || || || || || || ||
QY 891 QSCGRVSCSCSKSDAEPFTRCTKTVSGTSGVQSGPNSLSDICLOGNERPHLYEYS 950
  || || || || || || || || || || || || || || || || || || || || ||
Db *607 -----GKLLVNGSDA---IALKIKTILHFSAT-----NDR----- 633
  || || || || || || || || || || || || || || || || || || || || ||
QY 951 GDVQKQETTNVAQKPDPLEKTNWKSVCFGQPRNDTNWQTPTPSSSYEQC---ATQPHVL 1008
  || || || || || || || || || || || || || || || || || || || || ||
```

RESULT 14

T48818

glucan 1,4-alpha-glucosidase related protein [imported] - Neurospora crassa

N:Alternate names: protein 68B2.20

C:Species: Neurospora crassa

C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000

C:Accession: T48818

R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Partmann, B.; Nyakatu

submitted to the Protein Sequence Database, April 2000

A:Reference number: Z24541

A:Accession: T48818

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2022 <SCH>

A; Cross-references: EMBL:AL353821; GSPDB:GN00112; NCSP:68B2.20
A; Experimental source: cosmid contig 68B2; strain 74
C; Genetics:
A; Gene: NCSP:68B2.20
A; Map position: 2
A; Introns: 1192/3

Query Match	2.48;	Score 217;	DB 2;	Length 2022;
Best Local Similarity	17.28;	Pred. No. 0.0017;		
Matches 340;	Conservative 253;	Mismatches 720;	Indels 668;	Gaps 85;
QY	35	POOKPSKRKRKFMPKVVVVEGPKPKRPAELPKVVBVEGPKRKRPKKA-----ATOEKVK	89	
DB	4	PKPDPK-----PPSPRPKTPKTPTNKNGCAPTISPTKSATALGATTPKSV	50	
QY	90	SKETGSAKKNLKESATKK--PANVGD--MSNGSP-----EVLTKSCRKALNFLENPGD	140	
DB	51	PTPTNSGLSHHKPELKKPELSLLGDFLLGRPSPQORVAAQSAKSRKKTMSMDAQNVRE	110	
QY	141	ARQGDSESEIV--QNSGANSFSEIRDAIGCTNGSFLDSVSQDKTNGL-----GAMN	191	
DB	111	ELRQEMRAAAVRKLOQPGG-----VDRVKAWOKASQAQV---KAEGLPIPAEDARS	160	
QY	192	OPLEVSM---GNQPD-----KLSTGAKLARDOQDPLLFRNQOCQPPVATQNTQFPMENQO	243	
DB	161	EPTFAVNLSDGDVEEDRMKIKWRQPKPKKPKIET-----VPEKGT-----SGSD	208	
QY	244	AWLQMKNLQIGFPNGOQPRMTIRNQOQCLAMGNOQPMYLLGTPRPALVSGNQQLGGPQG	303	
DB	209	STAKEKNKVLG-RMADDRPDVILKEAR-----RP---GLPKKRIISDDHWMKKQO-	255	
QY	304	NKRPIFLNHQFLCPLAGNOLYCSPTDMHOLVMTSGGOHQHLLIKNQOCSLIRGOQPCVPL	363	
DB	256	-----KGKGLAKSP-----VPEAGQ-----PIPKDFLKRTA-----	281	
QY	364	IDQOPATPKGFTHLNQMVATSSWSGPLRPHSQSQVPTTYLHVESVSRILNGTTGTQCRSR	423	
DB	282	--QNPVSQKKIRDWAQRVLEPPPP--PVAARRPIVKIYRHAKT-----GETVIVEVDE	330	
QY	424	APAYDSLOQDTHQGNKYILSHSEINGNCKKALPONSLPTPIMAKLEAGSKRQYHRA	483	
DB	331	-----DDDHHRG-----ALSEPDM-----PRKRKSS	351	
QY	484	MGOTEKHDLNLAQIIAQSQD-----VERHNSSTCEVLDAAKTKIKVVOENLHGHP	537	
DB	352	SGESRRSSNPRTQCPNDGIRVRSLOKRGND-----DGIRISPVKPA-----RSLPD	401	
QY	538	EVITIEDDP---TDGARKGNKTASIS-----KGAS-----KGNSSPVKK	573	
DB	402	DGIKVRCPGPVYASDSSRSISITVSPSSSGRTGPSDRSGSRTPPRGASPPPRASIPRRA	461	
QY	574	TAEEKICIVKTPAKKGRAGKKSV-----PPPAHASEIQLMQPTPP	615	
DB	462	STPLRKASTPKPRASDHSAASDDVIEIVPESEVSSKRSPSPPKR-----RLRSPPPP	517	
QY	616	KTPLRSRK-----PKGCKGRKSIOD	634	
DB	518	KKLUPRRRSGGARRPKRRSPSPPTATQTTETTTDDRRPGADKPMPTPRNNGSGSD	577	
QY	635	SGKARGPSGELLQCDSTAEIITY---RMONLYLGDKEREQEQONAMVLYKGDGALVPYESKK	691	
DB	578	SDR-RPTAAGI--DGLAEIIPGFSASELELPLRGHTQRTHA-----RQS	620	
QY	692	RKRPKVDIDDETRINWLL-----MKGDEKGEDEBKOKKKEKWEEERRV	738	
DB	621	AKRPKQPNRESLKLVPNLVKVMVTGAMEKMQEMAEAPPPPTGKNKPARIESW-----	673	
QY	739	FRGRADSFIAHMLVQGDRRSPWKGSVVDSVIGVFLTQNVSDHLLSSAFMSLAARFPFK	798	
DB	674	LNTVDFPVEGMPNLPPVVAPEPLRVSTPE-----RNSKEKILVDRDLFAHREGRAPER	725	
QY	799	LSSSREDERNVRSVVVEDPEGCIILNLEIPSWQBKVQHP-----SDMEVSGV	845	

Db 1715 S 1715

RESULT 15
T05324
hypothetical protein F1C12.80 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999
R:Accession: T05324
R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.; De
ewes, H.W.; Mayer, K.F.X.; Schueller, C.
submitted to the Protein Sequence Database, April 1998
A:Reference number: Z15408
A:Accession: T05324
A:Molecule type: DNA
A:Residues: 1-1188 <BEV>
A:Cross-references: EMBL:AL022224
A:Experimental source: cultivar Columbia; BAC clone F1C12
C:Genetics:
A:Map position: 4
A:Introns: 305/3; 348/2; 975/2; 1091/3; 1141/1
A:Note: F1C12.80

Query Match 2.3%; Score 211; DB 2; Length 1188;
Best Local Similarity 17.4%; Pred. No. 0.0016;
Matches 244; Conservative 219; Mismatches 532; Indels 406; Gaps 55;

Qy 78 KPRKATQEKVKSGTSAKKKLNKESATKKPANVGMNKSPEVTLKSKALNF--- 133
Db 20 RPHERTQ---RDNTNNKKK-----NNNNVNIHVAPEKLNLFVRD 63

Qy 134 DLENFG---DARQGESEIVNGSANSFSEIRDAIGTNGSFLDSVQIDKTNGLAM 190
Db 64 HLENCVSVDVDDSIKAVECCSNKSSVSDHR--LSKTESRQDVP--SSSLGSD 118

Qy 191 NOPLEVMGNQDKLSTGAKLARDQPDLLTRNQOCQFPVATQNTQPMENQOAWLOMKN 250
Db 119 SDP---NSGOPENFSRNAASLSVQIWEART-----TOQPPSSNQSLIDRT 161

Qy 251 QLIGF-PFGNQPRWTIRNQ---QPLAMGNQPMYLGTPRALVSGNQGLGGPOGNKR 306
Db 162 SMGNSALNSESLSVSKESQMIQTEECNNEEBEIECAPPLESGEKDREGVVRMDI 221

Qy 307 PIFLNHQTCLPAGNQLYSGPTDMHQLVMSTGGQGLLIKNOQPGSLIRGOQCPVPLDQ 366
Db 222 IRKLSNDSETITNDNGSSGNDNSKEVQTEARSPPQVACSPR---IRGRQALADLLVQ 277

Qy 367 QPA-TPKGFTHLNMVATSMSSPGLRPHSQSQVPTTYLHVESVSRILNGTTGTCQSRAP 425
Db 278 MTRDRKDLACLRRHCVSKFT-----NRGRIQST-LRIRCYERCLAIQGRHRSKSTSA 330

Qy 426 AYDSLQDDIHQGNKYILSHEISNGCKKALPONSLPTPIAKLEEARGSKROYHRAMG 485
Db 331 GSDSNRRSGSGVHMLLRKY-----KANSNETGTASTSHASTGRIMDKDHPKA-- 381

Qy 486 QTEKHDLNLAQQAQSDQVVERHNSCTCYBLDAAKTK--IQKVVOENLHGMP-----EVI 540
Db 382 -TEKKVL-----QETIEKSVKEIKVSAVENAKKGVLSISENRLKLAGETI 432

Qy 541 ETEDDPTDGARKGNTASISKASKGNSFPVKTAKEKCIVPKTPAKKGRAGRKSVPP 600
Db 433 QKETVEGKRETKREKIVSKESLEKGE--KRESTKEKAIKESVAEKALVG----- 483

Qy 601 PAHASEIQWLQPTPPKTPLSRKPKGKRSIODSGKARG----- 640
Db 484 --IAEKVNLW-----NSDEKMRKVVVEKGTGENTINTERVNDVLEATRI 530

Qy 641 -----PSGELLQDSTAEIIRYMN-----LYLGDKEREQEQNAMYLYKGDGA 683
Db 531 LSVESAERSSTTTSTMTTRCEVAEKVVKKKKEDFVSLSNKSKSEVEDGNI--KPGQV 587

Qy 684 LVPYESKKRKRPRKVDIDDETTRIINLLMGKDEKEDEKOKKKKKEWEEERRVFRGA 743

Db 588 TTQADSCLRRKPAIEEKVLOETI-----VKSDLKKPTKESREEREIEEESTNGIA 640

Qy 744 DSFIARMHLVQGDREFPWKSGVVDVSVIGVLTQNVSDHLSASFMSLAARFPKLSRR 803

Db 641 EKVNLWSREKKRRKAMEKGR--GKTEGKAITE-----TNELLQEASR--RISNVE 688

Qy 804 EDERNVRSVVVDEPCECILNLEIPSWQKQVHPDSMEVSGVDSGSKQELRDCSNSGIER 863

Db 689 TAERSIDT-----SWITVVKVGRVIMDKRKKSGSETTSAEIGGGK 731

Qy 864 FNFLEKSIONLEEVLSODSFDPAIFQSGRVRGSCSCSDAEFPTTRCETKTVSGTSQ 923

Db 732 EEDL-ASVEAKSKDVIEDKNMNPQAVIH-----GS---KERKE-----RN 768

Qy 924 SVQTGSPNLSDEICLQGNRPHLYEGSGDVQKQETT-----NVAQKPDLEKTNMKD 976

Db 769 SSONGEETLS---LRNSEA---KSTKEIERQEVQTEEKSVSHGSKREKDKERNSLQYGE 820

Qy 977 SVCFGQPRNDTNWQTPSSSYEOCATRQPHVLDIEDFGQGEGLGYSWMSISPRVDRVN 1036

Db 821 KMCF--LRNS-----EAKSTKE-----IERNKS 841

Qy 1037 KNVPRFFRQGGVPRFTGQIIPSTPHLPGLSGSSSAVOEHODDTQHNOODEMNKA 1096

Db 842 QEV-----SQGEESASGHSRESAKENSKSQDD-----869

Qy 1097 SHLQTKFLDLLNSSECLTRQSTKONITDGLPRDRTAEDVVDPLSNSSLNQILVESN 1156

Db 870 -----ETSTHRN-----PNDK--KGIKEPEDE-----ESK 892

Qy 1157 SNKQETAVEYKETNATILREMKGT-----LADGKKTQSQWDLR 1196

Db 893 KYEREETGENVEEASVEFVNDWDGNEEBEBEYGDYFNGDDDDHIDISRRPSYWECLR 952

Qy 1197 KDVEGNEGROERNKNMDSIDYEAIRRASISEISEAIKERGMNMLAVRIKDFLERI--- 1253

Db 953 KE-----RYLEVINTSEKK-----DICNLIERTVSNFLTSLRQKIDNLMIT 996

Qy 1254 -VKDHGGIDLEWLRESPPDKAKDYLLSIRGLGLKSVCEVRLTLHLNLAAPPVDT---NVGR 1309

Db 997 RVQSHIGVPLNQIEEG-DEYEEWEVECSARNEEDNETEEPEKTNLEAPSDVCSOSSAR 1055

Qy 1310 IAVRMGW-----VPLQPLPESLQ--LHLELYPVLESIQKFLWPR 1347

Db 1056 SSTMSNWFRRDQDIDKDNPTTSLSLPELVPPTNQSTQDQTISDLKEQMEQLQREM--- 1112

Qy 1348 LCKLDORTL-----YELHVQ 1362

Db 1113 ---LELRNTVKSCIDMQLHFQ 1130

Search completed: July 5, 2001, 12:45:38
Job time: 200 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 12:42:18 ; Search time 25.55 Seconds
(without alignments)
5154.826 Million cell updates/sec

Title: PCT-US01-13059-2

Perfect score: 9089
Sequence: 1 MQSIMDSSAVNATEQND.....PRPLMARLHPASKLKNKT 1729

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_68:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3949	43.4	1017	2 T48452	hypothetical prote
2	2863	31.5	555	2 T48453	hypothetical prote
3	1559.5	17.2	1207	2 D84781	hypothetical prote
4	1260	13.9	234	2 T48454	hypothetical prote
5	709	7.8	917	2 T05430	hypothetical prote
6	278	3.1	5327	2 T13564	microtubule-associ
7	246.5	2.7	2897	2 B48666	cell proliferation
8	246.5	2.7	3256	2 A48666	cell proliferation
9	240.5	2.6	2938	2 T30249	hypothetical prote
10	233.5	2.6	2218	2 B4683	hypothetical prote
11	232	2.6	259	2 D75275	endonuclease III
12	219.5	2.4	1871	2 D79796	probable heat shoc
13	217.5	2.4	1490	2 T20513	hypothetical prote
14	217	2.4	2022	2 T48818	glucan 1,4-alpha-g
15	211	2.3	1188	2 T05324	hypothetical prote
16	210.5	2.3	1805	2 A34736	nestin - rat
17	208.5	2.3	3924	2 S37431	ankyrin 2, neurona
18	207	2.3	2331	2 T25410	hypothetical prote
19	206	2.3	3187	2 JC5837	364k Golgi complex
20	205.5	2.3	2562	2 T14266	xin protein - chic
21	205	2.3	1983	2 T00385	KIAA0624 protein -
22	204.5	2.2	2717	2 A34203	DNA-binding protei
23	197.5	2.2	1145	2 T18235	transcription acti
24	197	2.2	1366	2 B86292	hypothetical prote
25	196.5	2.2	3225	2 I52300	giantin - human
26	196	2.2	3488	2 T34418	hypothetical prote
27	192.5	2.1	2954	2 T14156	kinesin-related pr
28	192.5	2.1	3259	1 A56539	giantin - human
29	192	2.1	1888	2 T14273	zinc finger protei

30 191.5 2.1 2094 2 S33124
31 191 2.1 1742 2 T49451
32 190.5 2.1 2447 2 T16870
33 190.5 2.1 4957 2 T03455
34 190.5 2.1 5262 2 T03454
35 190 2.1 1284 1 NWVZAI
36 190 2.1 2464 1 QRMSP1
37 189.5 2.1 236 2 S75373
38 189.5 2.1 1634 2 T26517
39 189.5 2.1 1804 2 T34518
40 189.5 2.1 3498 2 T22330
41 189 2.1 2453 2 S60254
42 188.5 2.1 1435 2 A37793
43 188.5 2.1 1926 2 JC4842
44 188 2.1 1593 2 T22028
45 187.5 2.1 264 2 C72770

ALIGNMENTS

RESULT 1

T48452
hypothetical protein T32M21.160 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T48452

R:Bevan, M.; Terry, N.; Ardiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.;

ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.

submitted to the Protein Sequence Database, March 2000

A:Reference number: Z24487

A:Accession: T48452

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1017 <BEV>

A:Cross-references: EMBL:AL162875

A:Experimental source: cultivar Columbia; BAC clone T32M21

C:Genetics:

A:Map position: 5

A:Introns: 167/1; 874/1

A:Note: T32M21.160

Query Match 43.4%; Score 3949; DB 2; Length 1017;
Best Local Similarity 99.9%; Pred. No. 2.3e-206;
Matches 755; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQSIMDSSAVNATEQNDGSRQDVLEFDLNTPOQKPSKRRKMPKVVVEGPKRKP 60
DB 235 MQSIMDSSAVNATEQNDGSRQDVLEFDLNTPOQKPSKRRKMPKVVVEGPKRKP 294
QY 61 RKPAPLPKVVVEGPKRKPRAATQEKVSKETGSAKKNLKESATKKPANYGDMNSKSP 120
DB 295 RKPAPLPKVVVEGPKRKPRAATQEKVSKETGSAKKNLKESATKKPANYGDMNSKSP 354
QY 121 EYTLKSCRKALNFDLEPNQDARQDSEIIVNSGANSFSEIRDAIGTNGSFLDSVSQ 180
DB 355 EYTLKSCRKALNFDLEPNQDARQDSEIIVNSGANSFSEIRDAIGTNGSFLDSVSQ 414
QY 181 IDKTNGLGAMNPLEVSMGNQDPDKLSTGAKLARDQDPDLITRNQCCQFFVATQTFPME 240
DB 415 IDKTNGLGAMNPLEVSMGNQDPDKLSTGAKLARDQDPDLITRNQCCQFFVATQTFPME 474
QY 241 NQAWLQMKNLQIGFFPGNQPRMTIRNQOCLAMGNQOQPMYLLIGTPREALYSGNQLGG 300
DB 475 NQAWLQMKNLQIGFFPGNQPRMTIRNQOCLAMGNQOQPMYLLIGTPREALYSGNQLGG 534
QY 301 POGNKRPIFLNHTQCLPAGNQLYGSPTDMHOLVMTSGGQHQGLLKNQOQPSGLIRGQPC 360
DB 535 POGNKRPIFLNHTQCLPAGNQLYGSPTDMHOLVMTSGGQHQGLLKNQOQPSGLIRGQPC 594
QY 361 VPLIDQOQPATPKGFTHLNMVATSMSSPGLRPHSQSVPTTYLHVESVRIINGTTGTCQ 420

THIS PAGE BLANK (USPTO)

Db 595 VPLDQQPATPKGTHLNMVATSMSSPGLRPHSOSQVPTTYLHVSVSRILNGTGTGQC 654
Qy 421 RSRAPYDSLOQDHOQNKYILSHEISNGCKKALPONSSILPTPIMAKLEARGSKROY 480
Db 655 RSRAPYDSLOQDHOQNKYILSHEISNGCKKALPONSSILPTPIMAKLEARGSKROY 714
Qy 481 HRAMGQTEKHDNLNAQQAQTAQODVERHNSSTCEVYDAAKTKTKQVVQENLHGMPPEVI 540
Db 715 HRAMGQTEKHDNLNAQQAQTAQODVERHNSSTCEVYDAAKTKTKQVVQENLHGMPPEVI 774
Qy 541 EIEDDPTDARKGKNTASISGASKGNSPVKKTAEKEKCLVPTPAKKGAGRKKSVP 600
Db 775 EIEDDPTDARKGKNTASISGASKGNSPVKKTAEKEKCLVPTPAKKGAGRKKSVP 834
Qy 601 PAHASEIQLWPTPKTPLSRKPKGKRKSIQDSGKARGPSGELLCCDSTAEIIRYMON 660
Db 835 PAHASEIQLWPTPKTPLSRKPKGKRKSIQDSGKARGPSGELLCCDSTAEIIRYMON 894
Qy 661 LYLGDKEREQSONAMVLYKGDGALVYVESKKRPRKYVIDIDETTRIWNLLMGKDEKEG 720
Db 895 LYLGDKEREQSONAMVLYKGDGALVYVESKKRPRKYVIDIDETTRIWNLLMGKDEKEG 954
Qy 721 DEKDQKKEKWEERRVFRGRADSFIAHMLVQCD 756
Db 955 DEKDQKKEKWEERRVFRGRADSFIAHMLVQCD 990

RESULT 2
T48453
hypothetical protein T32M21.170 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48453
R:Bevan, M.; Terry, N.; Argiles, W.; Buysshaert, C.; Dasseville, R.; De Clerck, R.; De
ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, March 2000
A:Reference number: 224487
A:Accession: T48453
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-555 <BEV>
A:Cross-references: EMBL:AL162875
A:Experimental source: cultivar Columbia; BAC clone T32M21
C:Genetics:
A:Map position: 5
A:Introns: 469/3; 496/2; 524/3
A:Note: T32M21.170

Query Match 31.5%; Score 2863; DB 2; Length 555;
Best Local Similarity 100.0%; Pred. No. 7.8e-148;
Matches 549; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 795 SSAPFSLAARPPKLSRREDERNRVSVDPECCILNLEIPSWQKQVHPSPDMEVSG 844
Db 7 SSAPFSLAARPPKLSRREDERNRVSVDPECCILNLEIPSWQKQVHPSPDMEVSG 66
Qy 845 VDSGSKELRDCNSGIERFNFLEKSIONLEEVLSQDSFDPAIFQSGRGVSCSCSKS 904
Db 67 VDSGSKELRDCNSGIERFNFLEKSIONLEEVLSQDSFDPAIFQSGRGVSCSCSKS 126
Qy 905 DAEFPTTRCTKTVSGTSQSVQTPGNLSDEICLQGNRPHLYEGSDGVQKQETTNAQK 964
Db 127 DAEFPTTRCTKTVSGTSQSVQTPGNLSDEICLQGNRPHLYEGSDGVQKQETTNAQK 186
Qy 965 KPDLEKTNWQKSVCFQGPNDTNQWTPSSSYEQCATRQPHVLIDIEFGMQGELGYSW 1024
Db 187 KPDLEKTNWQKSVCFQGPNDTNQWTPSSSYEQCATRQPHVLIDIEFGMQGELGYSW 246
Qy 1025 MSISPRVDRVKNKVPFRFFRGGSVPREFTGQIIPSTPHLPGMGLSGSSSAVOEHODD 1084
Db 247 MSISPRVDRVKNKVPFRFFRGGSVPREFTGQIIPSTPHLPGMGLSGSSSAVOEHODD 306

Qy 1085 TOHQDEMKNKASHLOKTFDLNLSSEECITROSSTKONITDGLCPDRDRAEDVVDPLSN 1144
Db 307 TOHQDEMKNKASHLOKTFDLNLSSEECITROSSTKONITDGLCPDRDRAEDVVDPLSN 366
Qy 1145 NSSLQNLIVESNNSKEQTAVEYKETNATILREMKGTADGKPKTSOWDSLKRDVEGNEG 1204
Db 367 NSSLQNLIVESNNSKEQTAVEYKETNATILREMKGTADGKPKTSOWDSLKRDVEGNEG 426
Qy 1205 ROENKNNMDSIDYEAIRRASISEISAIEKRGNNMMLAVRIKDFLERIVKDHGIDLEW 1264
Db 427 ROENKNNMDSIDYEAIRRASISEISAIEKRGNNMMLAVRIKDFLERIVKDHGIDLEW 486
Qy 1265 LRESPPKAKDYLLSIRGLGKSVCEVRLTLHLNLAFFVDTNVGRIAVRMGWVPLQPLPE 1324
Db 487 LRESPPKAKDYLLSIRGLGKSVCEVRLTLHLNLAFFVDTNVGRIAVRMGWVPLQPLPE 546
Qy 1325 SLQLHLEL 1333
Db 547 SLQLHLEL 555

Query Match 17.2%; Score 1559.5; DB 2; Length 1207;
Best Local Similarity 28.5%; Pred. No. 1.2e-76;
Matches 480; Conservative 199; Mismatches 366; Indels 641; Gaps 61;
Qy 13 TEATEQNDGSRQDVLFDLANKTPQOKPSKRRKPKMPKVVEGPKPKRPAELPKVVE 72
Db 65 TEEVESLSSVSNVYAE-QILKTP-EKPKRKH--PKVREAKPKRPPKPRKSVVTD 120
Qy 73 GKPKRPRKAATQEKVK-SKETGSAKKNLKESATKKPANVGDMSNKSPEVTLKSCRKAL 131
Db 121 QGSKTPKRVKRVKVEVSKD-----QDAPVSSAAVETSTPK---RLCRVL 167
Qy 132 NFDLEN-----PGDARQ-GDSEIVONGSANSFSFEIRDAIGTNGSFLDSVQIDKTN 185
Db 168 DFEAENGENTNGDIRAGEMESALQEKDLSGN-QELKDCI-----LSAPSTPKRK 219
Qy 186 GLGAMNPLEVSMGNQDPK-----LSGAKLARDQDPDLLTRN-----QOCOPF 229
Db 220 SQGRK-----GVQPKKNGSNLEEVDISMAQAARQGTCCDMNLSGTYDEQCDY- 271
Qy 230 VATQNTQPMENQOAVLQMKNLGIFPFGNQOPRMTIRNQOQCLAMGNQOQPMYLGTPRP 289
Db 272 -----QKHWLYSPNL-----QOQGM--RYDAIC----- 293
Qy 290 ALVSGNQOLGQPGQKNRPIFLNHOTCLPAGNQLYSGPTDMHQLVMSTGGQOHLINQK 349
Db 294 -----SKVESQOQHNYSAF--HATCYSSTSQLSANRV-----LTVEERR 331
Qy 350 PGSILRQOPCVPLIDQOPATP-----KGFTLNQMVATSMSSPGLRPHSOSQV 398

THIS PAGE BLANK (USPTO)

Db 332 EGIFQGRQSENLVLSKIDTPIKKKTTGHARPNLSNMKLV-----EV 376
 QY 399 PTYYLHVSVSRILNGTTTCORSRAPAYDSLQODTHQNKVILSHISNGNGCKKALPO 458
 Db 377 PE---HL-----TSGYCSKPO-----QNNKILVDFRVI-----401
 QY 459 NSSLPPIMAKLEAARGSRQYHRAMGOTEKHDLNLAQOIAOSQDVERHNSSTCVYELDA 518
 Db 402 -----VSKKKPTKSEKQ-----414
 QY 519 AKKTKIOKVQENLHGMPEVIEIEDPTDGARKGKNTASISKGASKNGNSPPVKTKAEKE 578
 Db 415 ---TK-QKNLLPNLCRFPFSFGLSPD-----437
 QY 579 KCIYKTPAKKGRAGRKSVPPPAHASEIQLMQPTPKTPLSRKPKGKRKSIODSGKA 638
 Db 438 -----ELW-----KRRNSI-----446
 QY 639 RGPSELLCQDSTAEIIRYMNLYLGDKEQEONAMVLYK-----GDGALVPYE 688
 Db 447 -----ETISEL-----LRLLDINREHSETALVPTMNSQIVLFGGAGAIYPTV 490
 QY 689 S-KKKRPRKVDIDETRIWNLMLGKDEKGEDEKDKKKEWEEERRVFRGRADSFI 747
 Db 491 PVKKPRPKVLDDETRWKLLE-ENINSEGVDSDEQAKAKWEEERNVFRGRADSFI 549
 QY 748 ARMLVQGRDRFPKPGSVVDSVIGVFLTONVSHLSSAFMSLAARFP-PKLSSSREDE 806
 Db 550 ARMLVQGRDRFPKPGSVVDSVIGVFLTONVSHLSSAFMSLAARFP-PKLVSSNFD- 608
 QY 807 RNVRSVVDEPCCIUNLEIFSWQ-----EKVOHPSDMEVSGVDSKSEKOLRDCNS 859
 Db 609 -----AGTSMPSIQITYLDSSETMSPDHNHSSVT-----640
 QY 860 GIERFNFLKSIONLEEVLSODSDPDPAIFQCGRGVSCSKSDAEPPTTRCET--KT 917
 Db 641 -----LANTQDEKDVPSNET-----SRSSSEIAISAHESVDKT 676
 QY 918 VSCTS--QSVQTGSPNLSDEICLOGNERPHLYEGSGDVOK-QETTINVAOKKPDLEKTMN 974
 Db 677 TDSKEYVSDRKG-----SVEVDKTEKCRVNLNLFPSSEDSALTC 716
 QY 975 KUSVCGQPRNDTWGTTTSSVEQCATRQPHVLDIEDFGMQGEGLYGWSMSISPRVDRV 1034
 Db 717 QHSWSDAPQNT---BRAGSS-----EIDLEGE-YRTSFMKL-----750
 QY 1035 KKNVPRFRFGGSPVREFTGQIIPSTPHELPGMGLSGSSSAVQEHODDTQHNOQDEMN 1094
 Db 751 -----LOGVQVSLSDSNQVSPN-----NSPGDCSSEIKGFQ-----SMKEPT 787
 QY 1095 KASHLOKTFDLNLSSEE-CLTRQSTKQNTDGLPRDRTADVDVPLSNSSLNQILV 1153
 Db 788 KSS-----VDSSEPGCCSQ-----DG-----DVL-----807
 QY 1154 EGNSSNKEQTAVEYKTNATILREMGKTLADGKKPTSQWDSLRKDYEGNEGROERNKNNM 1213
 Db 808 ----SCQKPT---LKEGKKVLKE-----EKKAFDMDCLREAAQARAGIREKTRSTM 852
 QY 1214 DSDYDAIRASISSEISBAIKERGMNMLAVRI-----KDFLERIVKDHGGIDLE 1263
 Db 853 DTVDMKAIIRAADYKEVAETIKSRGMNHKLAERIQYTLNKMIMQGFDLRLVNDHGSIDLE 912
 QY 1264 WLRESPPDKAKYLLSIRGLGLKSEVCVRLTLNLNLAFPVDITNVGRIAVRMGWVLPQLPL 1323
 Db 913 WLKRVPPDKAK-----923
 QY 1324 ESLQLHLELYPVLESIOKFLWPRCLCKLDORTLYELHYQLITFGKVCCKSRPNCNACPM 1383
 Db 924 -----YELHYQMIFGKVCCKSKPNCNACPM 950
 QY 1384 RGCERHFASAYAGARLALPAPERSTTSATIPVPPSPFPVPAIMTELPLPLEKSLASGA 1443
 Db 951 KGCERHFASAFARLALPSTEGKMGMTPOKNPLP-----LHLPEPFQREGQSEV 999

QY 1444 PSNREN-----CEPIIEEPASPGQCEITEITSDIEDAYNEDPDDEIPTIKLINISQFGMT 1497
 Db 1000 VOHSEPAKKVTCCETPIIEEPASPEETAETAEVSIADIEEAFF-EDPEEPTIRLNDFAFTSN 1058
 QY 1498 LREHMERNELOEGDMSKALVALHPTTTSIPTPKLNISRLRTEHQVYELPDSHRLDGM 1557
 Db 1059 LAKIMEHNELOEGDMSKALVALHPTTTSIPTPKLNISRLRTEHQVYELPDSHRLDGM 1118
 QY 1558 DKREPDPSPYLLAIWTPGETANSAPPEQKCGKASGKMGCFDETCSECSLSREANSQV 1617
 Db 1119 EKREPDPSPYLLAIWTPGETANSAPPEQKCGKASGKMGCFDETCSECSLSREANSQV 1178
 QY 1618 RGTLLI 1623
 Db 1179 RGTILV 1184

RESULT 4
 T48454
 hypothetical protein T32M21.180 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
 C:Accession: T48454
 R:Bevan, M.; Terry, N.; Ardiles, W.; Buyssehaert, C.; Dasseville, R.; De Clerck, R.;
 ewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
 submitted to the Protein Sequence Database, March 2000
 A:Reference number: 224487
 A:Accession: T48454
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-234 <BEV>
 A:Cross-References: EMBL:AL162875
 A:Experimental source: cultivar Columbia; BAC clone T32M21
 C:Genetics:
 A:Map position: 5
 A:Introns: 49/2; 61/3; 81/1; 127/3; 149/3; 188/1; 201/1
 A>Note: T32M21.180

Query Match 13.9%; Score 1260; DB 2; Length 234;
 Best Local Similarity 100.0%; Pred. No. 2e-61;
 Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1496 MTLREHMERNELOEGDMSKALVALHPTTTSIPTPKLNISRLRTEHQVYELPDSHRLD 1555
 Db 1 MTLREHMERNELOEGDMSKALVALHPTTTSIPTPKLNISRLRTEHQVYELPDSHRLD 60
 QY 1556 GMDKREPDPSPYLLAIWTPGETANSAPPEQKCGKASGKMGCFDETCSECSLSREANSQ 1615
 Db 61 GMDKREPDPSPYLLAIWTPGETANSAPPEQKCGKASGKMGCFDETCSECSLSREANSQ 120
 QY 1616 TVRGTLILPCRTAMRGSPFLNGTYFOVNEFLADHESLKPIDVPRDWIWDLPRTYVFGT 1675
 Db 121 TVRGTLILPCRTAMRGSPFLNGTYFOVNEFLADHESLKPIDVPRDWIWDLPRTYVFGT 180
 QY 1676 SVTISIFRGLSTEQIQCFWKGVFCVGFGEQKTRAPRLMARLHFFPASKLNKNT 1729
 Db 181 SVTISIFRGLSTEQIQCFWKGVFCVGFGEQKTRAPRLMARLHFFPASKLNKNT 234

RESULT 5
 T05430
 hypothetical protein F28A23.180 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 24-Nov-1999
 C:Accession: T05430
 R:Bevan, M.; Weichselgartner, M.; Fartmann, B.; Granderath, K.; Dauner, D.; Herzl, A.
 submitted to the Protein Sequence Database, October 1998
 A:Reference number: Z15415
 A:Accession: T05430
 A:Molecule type: DNA
 A:Residues: 1-917 <BEV>

THIS PAGE BLANK (USPTO)

A;Cross-references: EMBL:AL021961
 A;Experimental source: cultivar Columbia; BAC clone F28A23
 C;Genetics:
 A;Map position: 4
 A;Introns: 41/3; 70/3; 88/1; 118/1; 321/1; 367/1; 474/1; 504/3; 655/2; 667/3; 695
 A;Note: F28A23.180
 C;Superfamily: Arabidopsis thaliana hypothetical protein F28A23.180

Query Match 7.8; Score 709; DB 2; Length 917;
 Best Local Similarity 21.4; Pred. No. 1.1e-30;
 Matches 232; Conservative 108; Mismatches 163; Indels 580; Gaps 22;

Qy 651 IAEIIRMONLYLGDKEREQONAMVLYGKDALVPEYSEKRRKPRKPDIDETRIWNL 710
 Db 398 IAKLIKDMGRKLNK-----VTTMIKADKLV-----TAKYNLOPETIKENDV 431

Qy 711 LMKGDEKEDEKDEKKEKWEERRVFRGRADSFIAHMLVQDRFSPWKGVSDSV 770
 Db 432 LM--VNDSPRSYDDKETEAKKKEREITFQTRIDLFINRHRLOGNRKFKQWGVSDSV 489

Qy 771 IGVFTONVSHLSSAFMSLAAREPPKLSRREDERNVRSVVVEDPESGILNLEIPSW 830
 Db 490 VGVFTQNTDYLSSNAGMSVAARPP-----VDAREGLSYIEEPQ----- 530

Qy 831 QEKVQHPDMEVSGVDSGSKQLRDCNSGIERFNFLEKSIQNLFEELSSQDSFPAIF 890
 Db 531 -----DAKSECI----- 538

Qy 891 QSGRGVSCSKSDAEPTTRCETKTVSGTSQSVQTSQSPNLSDICLOGNERPHLYEGS 950
 Db 539 -----ILSDP----- 543

Qy 951 GDVQKQETTNVAQKPDLEKTMNKNDSVCFGPRNDTNQWTPSSSYEQCATRQPHVLDI 1010
 Db 544 ----- 543

Qy 1011 EDFGMQGEGLGYSWMSISPRVDRVKNKNVPRFRFQGGVSPREFTGQIIPSTPHELPGM 1070
 Db 544 ----- 543

Qy 1071 LSGSSAVQEHODTQHNQODEMKNASHLQKTFDLNLSSECLTRQSTQNTQDGLP 1130
 Db 544 ---SISKVEDHEN-----TAKRKNKGTII-- 565

Qy 1131 RDRTAEDVVDPLSNSSQLNILVESNSSNKEQTAVEYKRETNATILREMGTLADGKKPTS 1190
 Db 566 ---EDEIVD----- 571

Qy 1191 QWDSLRKDVQEGNEGRQERNKNMDSIDYEAIRASISEISEAIKERGMNMLAVRIKDFL 1250
 Db 572 -WNLR--MYTKEG--SRPEMHDSVNSDVRLSGQNVLETTIKRGQFRILS----- 620

Qy 1251 ERIVKHGIDLEWLRSPDKAKDYLLSIRGLGKSVCEVRLTLHLNLAFPVDNNGRI 1310
 Db 621 ERIL-----VDNNGRI 632

Qy 1311 AVRGMVPLPPLPSIQLHLLLYPVLESIQKFLWRLCKLDQRTLYELHYLITFGKVF 1370
 Db 633 AVRGLVPLPLELPNGVQMH-----QLFEYELHYQMITFGKVF 669

Qy 1371 CTKSPNCNACPMRCEHRFASAYASARLALPAPERSLTSATIPVPPESF-----PPVA 1425
 Db 670 CTKTTPNCNACPMKCEKYFASAYVSSKVLLESPEKM-----HEPNTFMAHSQDVA 722

Qy 1426 IPMIELPLFLEKSLASGAPSNRENCEPIIEEPASPGQCEITEESDIEDAY---NEDPD 1482
 Db 723 VDMTSNINLVECVSSGCGDQAICYKPLVEFPSSPRAEIPEST--DIEDVPMNLYQSYA 780

Qy 1483 EIPTTKLINEOFGMTLRHM--ERNNELOEGDMSKALVALHPTTTSIPTP---KLNLSR 1537
 Db 781 SVPKTFDFDLDAKSVEDALVISGRMSSDSEISKALVPIPTPENACIPIKPPRKKMYNR 840

Qy 1538 LRTEHQVYELPDSHRLDGMKRPDDPSYLLAIWTPGETANSAPPEQKCGKASGKM 1597
 Db 841 LRTEHVYVLPDHNHLLDFERRKLDPSYLLAIWTP----- 878

Qy 1598 CFDETCSECSLRANSQTVRGVLLIPCRITAMRGSFPLNGTYFQVNFELFADHESLAPID 1657
 Db 879 ----- 878

Qy 1658 VPRDWINDLPRTVYFCTSVTSIFRGLSTEQIOCFWKFVGVCRGPFQKTRAPRLMARL 1717
 Db 879 -----GFLCLRAFDRKQDRPKELVRL 900

Qy 1718 HFP 1720
 Db 901 HTP 903

RESULT 6
 T13564
 microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
 N;Alternate names: hypothetical protein EG:49E4.1
 C;Species: Drosophila melanogaster
 C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
 C;Accession: T13564
 R;Spanos, L.; Papagiannakis, G.; Sidenkiamos, I.; Louis, C.
 submitted to the EMBL Data Library, April 1999
 A;Description: Sequencing the distal X chromosome of Drosophila melanogaster.
 A;Reference number: Z17689
 A;Accession: T13564
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-5327, <SPA>
 A;Cross-references: EMBL:AL0311128; PIDN:CAA20006.1
 C;Genetics:
 A;Cross-references: FlyBase:FBgn0025392
 A;Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
 A;Note: EG:49E4.1
 C;Superfamily: Drosophila 576K microtubule-associated protein homolog

Query Match 3.1; Score 278; DB 2; Length 5327;
 Best Local Similarity 19.7; Pred. No. 3.3e-06;
 Matches 277; Conservative 198; Mismatches 583; Indels 348; Gaps 59;

Qy 2 QSIMDSAVNATEQNDGSRQDVLEFDLNLKTPQQRPSKRKKFMPKVVEG-----K 55
 Db 1986 ESIKD-----EAKSKEESRRE-----SVAESPLPSKEASRPASVAESIKDEAEK 2031

Qy 56 PKRPRPAELPKVVVVGKPKRKP-----RKAATQEKVKSKETGSAKKNNKESATKK 108
 Db 2032 SKEESRRESVAESPLPSKEASRPASVAESIKDEAEKSKESRESAASPLPSKEASR 2091

Qy 109 PANYG---DMSNKSPEVTLKSCRKALNFLENPGDAR--QGDSE--SEIVONSSGANSF 160
 Db 2092 PASVAESVKDEADKSKESRRE-----SMAESGKAQSIKGDQSPKLEKSRPESVAESV 2144

Qy 161 SEIRDAIGGTGSLDSVSQIDKTNGLGAMNQPLEVSGMGNQPKLSTGAKLARDQQPDLL 220
 Db 2145 KD--DPVKSKEPSRRESVAGSVTADSDQSPLESKGASRPESVDSVDKAEKQESRR 2202

Qy 221 TRNQOCOFFVATQNTQPMENQQAQWLQKMLQIGFPFGNQOQPRWTIRNQOFCPLANGNOQ 280
 Db 2203 ESKTESVTPPKAKDKSPKEVLQ-----PVSMT-----TIREDA-----DQP 2240

Qy 281 MYLTGTRPALVSNQO-----LGGPQGNKRPIFLNHQTCPLPAGNQLYGSPTDMHQLV 333
 Db 2241 M-----KPSQAESRRRESIAESIKASPRDEKSPKASKEASRPESVAESIKYDLKPKQII 2294

Qy 334 MSTGQOQH---GLLIKN-----QOPGSLRQOQPCVPLID--QOPATPK-----GF 374
 Db 2295 KDDKSTESHRRESLEDKSAVTSEKSRPLSVASDHAEAAVAIEDDAKSSISPDKSRPGF 2354

Qy 375 THLNQMVATSMSSPGLRPHSQSQVPTTYLHVESVSR-----ILNGTTGTQCRSRAPAYD 428

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 12:42:17 ; Search time 31.78 Seconds
(without alignments)
3298.265 Million cell updates/sec

Title: PCT-US01-13059-2
Perfect score: 9089
Sequence: 1 MQSIMDSSAVNATEEQND.....PRPLMARLHPASKLKNKT 1729

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues
Total number of hits satisfying chosen parameters: 412676
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :			
A_Geneseq_0601.*			
1:	/SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*		
2:	/SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*		
3:	/SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*		
4:	/SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*		
5:	/SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*		
6:	/SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*		
7:	/SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*		
8:	/SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*		
9:	/SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*		
10:	/SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*		
11:	/SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*		
12:	/SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*		
13:	/SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*		
14:	/SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*		
15:	/SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*		
16:	/SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*		
17:	/SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*		
18:	/SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*		
19:	/SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*		
20:	/SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*		
21:	/SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*		
22:	/SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	246.5	2.7	3256	21 AAY50976	Human cell cycle p
2	237	2.6	2819	22 AAB35408	Human 07G27 gene
3	210.5	2.3	1805	13 AAR27204	Rat nestin. Rattu
4	210.5	2.3	1805	15 AAR60126	Rat nestin protein
5	208.5	2.3	2432	21 AAY85565	Human homologue of
6	195	2.1	1780	19 AAW53863	Human gravin polyp
7	195	2.1	1780	21 AAB15380	Human gravin prote
8	192.5	2.1	2954	20 AAY01632	Amino acid sequenc
9	191.5	2.1	1786	14 AAR41043	CD4-EBA175 fusion
10	190	2.1	1284	9 AAR81187	Sequence encoded b
11	189	2.1	2453	21 AAB12454	HNRCR protein sequ

12	188.5	2.1	1435	16 AAR70232	P. falciparum SABP
13	188.5	2.1	1435	18 AAW22477	Silic acid bindin
14	188.5	2.1	1435	21 AAY77900	P. falciparum SABP
15	188.5	2.1	1596	18 AAW31347	Rat tumour suppres
16	188.5	2.1	1604	16 AAR70105	TNF-R-EBA 175 fusi
17	187.5	2.1	1863	21 AAY56028	Human BRCA1 tumour
18	185.5	2.0	1312	21 AAY78884	Retinoblastoma bin
19	185.5	2.0	2409	12 AAR12609	Versican. Homo sa
20	185	2.0	2441	19 AAW40058	Cellular transcrip
21	185	2.0	2441	21 AAY94252	Mouse nuclear CREB
22	184.5	2.0	1346	18 AAW31346	Rat tumour suppres
23	184	2.0	3266	21 AAB42491	Human OREX ORF2255
24	183	2.0	2441	16 AAR79054	CREB binding prote
25	182	2.0	1863	18 AAW26522	Human BRCA1 consen
26	182	2.0	1863	19 AAW79665	BRCA1 (Omi1) prote
27	182	2.0	1863	19 AAW76100	Human BRCA1 omi3 p
28	182	2.0	1863	19 AAW76098	Human BRCA1 omi1 p
29	182	2.0	1863	21 AAB24217	Human BRCA1 (Omi1)
30	182	2.0	1863	21 AAB24218	Human BRCA1 (Omi2)
31	181	2.0	1931	21 AAB23258	Human apoptosis in
32	181	2.0	2781	21 AAY57453	Human transcriptio
33	180	2.0	2907	21 AAY57452	Human transcriptio
34	179.5	2.0	1863	18 AAW23286	Human breast and o
35	179	2.0	1619	17 AAR81528	BRCA1 mutant from
36	179	2.0	1663	15 AAR46608	Plasmodium falcipa
37	178.5	2.0	778	21 AAG46504	Arabidopsis thalia
38	177.5	2.0	3248	17 AAR99795	Kinetochore protei
39	177	1.9	1863	21 AAY77818	BRCA1 protein sequ
40	176.5	1.9	1484	20 AAW89721	Canine ribosome re
41	176.5	1.9	1978	20 AAY07032	Breast cancer asso
42	175	1.9	1210	16 AAR66450	AF-4 protein (enco
43	174.5	1.9	2101	21 AAY49936	Human NuMA protein
44	174	1.9	1863	17 AAR81536	BRCA1 mutant from
45	173	1.9	1863	16 AAR76641	BRCA1 protein. Ho

ALIGNMENTS

RESULT	1
AA50976	
ID	AA50976 standard; Protein; 3256 AA.
XX	
AC	AA50976;
XX	
DT	10-MAR-2000 (first entry)
XX	
DE	Human cell cycle protein Ki-67.
XX	
KW	Cell cycle protein; Ki-67; therapy; cell proliferation; allergy;
KW	tumor treatment; autoimmune disease; scar formation; inflammation;
KW	rheumatic disease; transplantation.
XX	
OS	Homo sapiens.
XX	
PN	DE19822954-A1.
XX	
PD	25-NOV-1999.
XX	
PF	22-MAY-1998; 98DE-1022954.
XX	
PR	22-MAY-1998; 98DE-1022954.
XX	
PA	(BORS-) FORSCHUNGSZENTRUM BORSTEL ZENT MEDIZIN.
XX	
PI	Flad H, Gerdas J, Boehle A, Deinert I;
XX	
DR	WPI; 2000-039964/04.
DR	N-PSDB; AAZ43872.
XX	
PT	Ki-67 gene antisense oligonucleotide
XX	
PS	Disclosure; Page 19-29; 36pp; German.

Db f645 vda1sqf1nnyasvviid---dhpevtviedpqsng---lnddgftevsvskkq----- 1691

Qy 1196 RKDVNGENGRGER-----NKNNDMSIDYAIRASISEATKRGKGMNMLAVRIKD 1248

Db 1692 qrlqdeerrkkeeqvgnkknane-----kgrsqtskl----- 1727

Qy 1249 FLERIVKDHGGIDLEWLRSPDPKAKDYLLSIRGIGLKSVECVRLTLTHNLAFPPVDITNNG 1308

Db 1728 -----pprfakkqatgiq---qasqasvvpplaspaplpstas 1763

Qy 1309 RIAVRMGWVP--LQPLPSLQHLLELYPVLESTQKFLWPRKLCDQRTLYELHYQLITF 1366

Db 1764 vpasaplattpvpastsa-----vpas-----t1 1792

Qy 1367 GKVFTCKRPNCNCPMRGCRHFASASARLALPAPERSLTSATIPVPESFP----- 1422

Db 1793 apvlstasap-vpaspi-----apvasasvsvpasasaaaitssasapasapapt 1844

Qy 1423 -----EVAIPMI---ELPLPLEKSLAS-GAPSNRENCEPIIEPPASPGOEITEESD 1471

Db 1845 llasvstpasvtilasasipi-lasalastapt---apaasspaap-----vitapt 1894

Qy 1472 TEDAYNEDPEIPTIKLNIEQFGMTLREHMERNMELQEGDMSKALVALHPTTTSIPTPK 1531

Db 1895 i-----pasapt-----asvplapasasapapa 1917

Qy 1532 LKNISRLRTEHOVELPDSHRLDGMKREPDPDPSPYLLAIWTGCTANSAQPPKCGG 1591

Db 1918 ptpvs-----apnpappap-----agtqaqthkpvgnplq----- 1947

Qy 1592 KASGKMCDFETCSECN-----SLREANSQTVRGTLIPCRMTAMRGSPFLNGTYFQVNELF 1646

Db 1948 -----ttsqskqpppsiripsaqtpngt----- 1971

Qy 1647 ADHESLKPIDVPRDW-----IWD 1665

Db 1972 -dyvasgksiqtpqshgtltaelwd 1995

RESULT 3

AAR27204

ID AAR27204 standard; Protein; 1805 AA.

XX AC AAR27204;

XX DT 20-MAY-1998 (first entry)

XX DE Rat nestin.

XX KW Intermediate filament; central nervous system; brain tumour;

XX KW neurofilament.

XX OS Rattus norvegicus.

XX PN W09214821-A.

XX PD 03-SEP-1992.

XX PF 21-FEB-1992; 92WO-US01375.

XX PR 22-FEB-1991; 91US-0660412.

XX PA (MASI) MASSACHUSETTS INST TECHNOLOGY.

XX PI Lendahl U, McKay RDG;

XX DR WPI; 1992-316175/38.

XX DR N-PSDB; AAQ28398.

XX PT Diagnosis of pre-disposition to brain tumours - using DNA

XX PT encoding nestin which distinguishes neural multi-potential stem

XX PT cells from neuronal, glial and muscle cells

PS Claim 3; Fig 1; 63pp; English.

XX CC The amino acid sequence of the nestin gene which was deduced from

CC the nucleotide sequence, suggests that nestin is a member of the

CC intermediate filament protein family. The rat nestin amino acid

CC sequence shows 75% similarity with the human nestin sequence. There

CC is more than 60% identity between the two sequences. Antibodies to

CC nestin protein can be used in vivo diagnosis of brain tumours.

CC See also AAQ28399 for the human nestin gene.

XX SQ Sequence 1805 AA;

Query Match 2.3%; Score 210.5; DB 13; Length 1805;

Best Local Similarity 18.7%; Pred. No. 7.6e-07;

Matches 283; Conservative 216; Mismatches 564; Indels 447; Gaps 72;

Qy 60 PRKPAELPVVVEGPKRKRKAATQEKVSK--ETGSAKKKLNKESATKKPANVGDMSN 117

Db 159 prrp-----papphripgapevedlarlrgvewrgavrdygervahmesslgqare 210

Qy 118 KSPVET--LKSCRKALNFDLENPGDARGDSESEIVQNSSGANSFSEIRDAI--GGTNGSF 174

Db 211 rlsqavrgarecr-----levqqiqadrsiqerrealqrlegw 251

Qy 175 LDSYSQIDKTN--GLGAMNOPLEVSMGNQPDKLSTGAKLARDQPDLLTRNQCCQFPVATQ 233

Db 252 qdrilatdkfqlavealeqekqqlsqiaqileggqqlah-----lkmnslevaty 303

Qy 234 NTQPFMENQAWLQMKNLQIGFPFGNQOQPRMTIRNQOQCLAMNQOQPMYLIGTRPPALVS 293

Db 304 rtlleaensr--lqtpgrgsqaslgfidpkl-----kpnf-lgipedyll-gsvlpalsp 354

Qy 294 GNOQLGGPQGNKRI--FLNHOTCLPAGNQLYGSPTDMHQLVMSTGGQOHLKLNQOQPG 351

Db 355 tsfsspplntletpvtafiktqeflqartptlas--tpippiseap-----cpnn 402

Qy 352 SLIRGQOQPCVPLIDQOQATP---KGFTLNQWVATSMSSPG-----LR 391

Db 403 aevraqevplsalqtqapeplwlkatvpsasailpeleepgkqgghfddltlatnln 462

Qy 392 PHSQSVPTTYLHVESVSRIILGTTCORSRAPAYDSLOODIHQNKYI-----LSHEI 446

Db 463 phh-----pt-----leakdgessesrvasi--fgedegqiwelvekeadvkv 505

Qy 447 SNGNGCKKALPONSLLPTPIMAKLEEARGS---KROYHRAMGOTEKHDLLAQIAQSQ 502

Db 506 ensaqk---tqesgldt-----eetqdsqgplqketikalg--eeplmslklqnyeta 554

Qy 503 DVERHNST-----CYEYLDAAKTKIKQVVQENL-----HGMP--PEVIEIDDDPTD 548

Db 555 gkencnssteghlgtlegpekeqkqiplksleeknvesektlengvvpvlsellgkedtrte 614

Qy 549 -----GARK-----GKNTASISKAGSKN-----SSPVKKT 574

Db 615 dqelmspgkgtlkrfssigkesqevrvpskegnleswtafkeeshp1gfpgaedqmierl 674

Qy 575 AEKEKICVPTKPAKKG---RAGRKKSVPPPAH-----ASEIQLMQPTPK 616

Db 675 vekedqfprspeeedqacrp1qkenqep1gyeaaegq1lerliekesqeslsrpeeed 734

Qy 617 TPLSRSKPKGK---GRKSTQDSGKARGPSGELLCQDSIAEIIYRMQNLVIGDK--EREQE 671

Db 735 qeagrs1qkenqep1gyeaaedqmler-----liekesqeskspeenqr1gkplerenq 789

Qy 672 QNAMVLYKGDGALVPYSEKSKRKRPRKVDIDDETRINWLLMGKDEKED--EEKDKKKEK 730

Db 790 kslryleengetfvplesrnqr1rsleveeeedqr1kplekvsqds1gs1aeenvqplr 849

Qy 731 WVEEE-----RRVFRGRADSF1ARMHLVQGDRRFRFPWKGVSVVDSVIGVFL-----TQNVS 780

Db 850 yleeddckinkslledkthks1gs1edrngds1iip-----qesetqvs1rpeeedqriv 904

```

QY 781 DHLSSAFMSLAARPPPKLSSSREDERNV--RSVVVEDEPGCILNLNETPSPQEKVQHP 838
Db 905 nhlekes-----qefsrseeeqymerslegenhe-----slssvekedqmvve--s 949
QY 839 DMEVSVDSG-----SKEQLRDCNSNGIERFNFLEKSIQNL-----EE 876
Db 950 qlekesqsgskledesqetfoplekenaeslrslagqdeeqkleqetqqltravgneq 1009
QY 877 EYLSQSDSPD-----AIFOSCGRVGSCS-CSKSDAEFPFTR-CETKTVSGTS-- 922
Db 1010 mavsppekvdelpkplngndqeiarslgkendesvlslekgietvkslelepleta 1069
QY 923 -----QSVQFGPNLSDEICLQNERPHLYE-----GSGDVQKQETTNVAOKKPDLEKT 971
Db 1070 eedlerrksidteplwstevaretvepedpgslgsldenretltslekesqeslsl 1129
QY 972 MNW-----KDSVCFQGNRDNFN-----WQTPSSSYEQ----- 999
Db 1130 gkwnvetrvdsqqclqveegllqeeqhqsleqvleqelpssgnqgrwedvvegkavgqea 1189
QY 1000 -CATRQPHVLDIEDFGMQGEG-----LGYSWM--SISPRVDRVKNKNVPR 1041
Db 1190 platgvgtedkaehhrggggeeeaaegellqdivgeawslgssepkeqrvpaealdn 1249
QY 1042 RFRGCGSVPREFTGQIIPSTHELPGMGLSCSSAVQEHQDD-TQHNQOQDMNKAHLQ 1100
Db 1250 ---leggal-----evp---vadsmpevterdedraqagqedsievtlgile 1289
QY 1101 --KTFLDLNLSSECLTROSSTKQNTDCLPRDRTAEDVDVPLSNSSSLONILVESNSS 1158
Db 1290 aartglel-----eqevvgledprhfareaaipsigeesvk----- 1326
QY 1159 NKEQTAVEYKETNATILRMKGTLDAGKKTPTSQWDSLRKDVEG-NEGROKNNKMNDSID 1217
Db 1327 -----akiaggleg---pgkep-----keagaldsgilelpktsseale 1362
QY 1218 YEAIRRASISEIAKERGMNMLAVRIKFLERIVKDHGGIDLEWLRESPPDKAKD-- 1275
Db 1363 cghhee---sesmegweeeas-----le--tsdhegsdapqpr--ppteedeeg 1405
QY 1276 YLLSIRGLGLKSVCEVRLTLHLNLAFFVDNNGRIAVRMGWVPLQPLPSQLHLELYP 1335
Db 1406 aqaaltapapkile-----pcspipiltadah--elqp 1435
QY 1336 VLESTQKFLW 1345
Db 1436 qaegiqeagw 1445

RESULT 4
AAR60126
ID AAR60126 standard; Protein; 1805 AA.
XX
AC AAR60126;
XX
DT 21-MAR-1995 (first entry)
XX
DE Rat nestin protein is useful to identify brain tumours.
XX
KW nestin gene; brain tumour; neoplastic cells; glial; neuronal;
KW muscle; neural multipotential stem cell; mammalian brain; detection;
KW diagnosis; medulloblastoma; glioblastoma; oligodendroglioma.
XX
OS Rattus rattus.
XX
FH Key Location/Qualifiers
FT Misc-difference 167..168
FT /note="encoded by CACGG"
FT Misc-difference 172..173
FT /note="encoded by CCG G GCC"
XX
PN US5338839-A.
XX

```

```

PD 16-AUG-1994.
XX
PF 12-APR-1988; 88US-0180548.
XX
PR 12-APR-1988; 88US-0180548.
PR 02-JUN-1988; 88US-0201762.
PR 25-OCT-1990; 90US-0603803.
PR 22-FEB-1991; 91US-0660412.
PR 19-MAR-1992; 92US-0853913.
XX
PA (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX
PI Lendahl U, McKay RDG;
XX
DR WPI; 1994-263332/32.
XX
N-PSDB; AAR70447.
XX
PT Nucleotide and protein sequences for human and rat nestin -
PT distinguishes neural multipotential stem cells and brain tumour
PT cells from more differentiated cell types; for use in the
PT diagnosis of brain tumours
XX
PS Claim 4; Column 35-44; 45pp; English.
XX
CC This sequence is the rat nestin protein encoded by AAR70447. Its
CC degree of sequence homology to intermediate filaments in the core
CC domain, its dissimilarity in the head and tail domains, and its
CC different splicing pattern suggest that it defines a new class of
CC intermediate filament protein. Nestin protein expression distinguishes
CC neural multipotential stem cells and brain tumour cells from the more
CC differentiated neural cell types (eg., neuronal, glial and muscle
CC cells of the adult brain). The nestin protein can be used in diagnosing
CC tumours of the brain, such as medulloblastomas, glioblastomas and
CC oligodendroglioma. (See also AAR60127).
XX
SQ Sequence 1805 AA;

```

```

Query Match 2.3%; Score 210.5; DB 15; Length 1805;
Best Local Similarity 18.7%; Pred. NO. 7.6e-07;
Matches 283; Conservative 216; Mismatches 564; Indels 447; Gaps 72;

QY 60 PRKPAELPVVEGPKRKPRKAAQTEKYKSK-ETGSAKKKNLKESATKKPANNVGDMSN 117
Db 159 prrp-----papphrpqpapevediarrigevrvavrdyqervahmesslgqare 210
QY 118 KSPVET--LKSCRKALNFDLENPGDARQDSESEIVQNSSGANSFSEIRDAI-GFTNGSF 174
Db 211 risqavrgarecr-----levqqlqadrdslqerrealqrlegrrw 251
QY 175 LDSVSQIDKTN-GLGAMNQPLEVSMGNQDPDKLSTGAKLARDQDPDLLTRNQOCQPPVATO 233
Db 252 qdrldatdkflavealeqekqglsglaileggqqlah-----lkmslslevaty 303
QY 234 NTQFPNENQOAWLQMKNLQIGFPFGNQQRMTIRNQOCLAMGNQOPMYLIGTPPALVS 293
Db 304 rtileaeansr--lqtprgsgaslgfldpki-----kpnf-igipedqyl-gsvlpalsp 354
QY 294 GNOQLGGPGQGNKRPI--FLNHQTCPLPAGNQLYCGSPDMHQLVMSTGGQGHLLIRNQOPG 351
Db 355 tsfpplntletpvtafiktqefiqartptlas-tpippiseap-----cpgn 402
QY 352 SIIRGOQPCVPLIDQOAPATP---KGFTHLNQMVATSMSSPG-----LR 391
Db 403 aevraqevplslqtqaepelwkatvpssailpelleepgkqgghfpdditslatnln 462
QY 392 PHSQSQVPTTYLHVESVSRILNGTTGTCORSRAPAVDSLQODIHQGNKYI-----LSHEI 446
Db 463 phh---pt-----leakdgesesrvssi--fgedegqlweiveheadlevkv 505
QY 447 SNGNGCKKALPONSSILPTIMAKLEBARGS----KROYHRAMGQTEKHDLNLAQQAQ 502
Db 506 ensaqaq---tqesgltdt-----eetqdsqqlqketikalg--eepimskiqnyeta 554

```

Qy 503 DVERHNSST-----CVEYLDAAKTKIKQVVOENL-----HGMP--PEVIEIEDPTD 548
 Db 555 gkencnsteghlgtieqpekekipiksleeknevsekltengvpsvlsellgkeditrte 614
 Qy 549 -----GARK-----GKNTASISKASKGN-----SSPVKKT 574
 Db 615 dqelmspgkgtlkrfsslgkesqevvrpskegnleswtafkeesqhpigfpgaedqmierl 674
 Qy 575 AEKEKIVPKTPAKKG---RAGRKKSVPPPAH-----ASEQLMQPPTPPK 616
 Db 675 vkeedqsfprspeeadeqarplkqenqeplygeaeagqilerliekesqeslrsppeed 734
 Qy 617 TPLSRKPKGK---GRKSIDSKARGPSGELLCQDSIABIIRYRMNLYLGDK-EREOE 671
 Db 735 qeagrsqkqenqelygeaeaqmle---liekesqeslkspeenqrigkpleranq 789
 Qy 672 QNMYLYKGDGALVYPSKSKKPKPKVDIDDETRINWLLMGKDEKEDG-EEKDKKKEK 730
 Db 790 kslryleenqetfvplesnqrprlsrleeeeeeqrivkplekvsqdsigslaeenvgplr 849
 Qy 731 WVEEE---RVFRGRADSFARMHLVQGRFRFPKWSVVDSVIGVFL-----TQNV 780
 Db 850 ylesddcinkslleedkthkslgslednrgdsliip-----gesetqvsrlrpeeedqrv 904
 Qy 781 DHLSSAFMSLAARFPKLSRSSREDERNV--RSVVVEDPEGICILNLEIPSWQEKVQHP 838
 Db 905 nhlekes-----qefrsseeeeqmierslegenhe-----slsvekedqmev-s 949
 Qy 839 DMEVSGVDVG-----SKEQLRDCNSGIERFNFLKSTQNL-----EE 876
 Db 950 qlekesqgksledesqetfglekenaeslrsagldqeeqlqetqqltravneq 1009
 Qy 877 EVLSQDSFDP-----AIFQSCRGVSCS-CKSDAEFFTR-CETKTVSGTS-- 922
 Db 1010 mavspkevdpelpkplngndqelarslgkenqeslvslekgietkvsleleiepleta 1069
 Qy 923 -----QSVOTGPNLSDETLCLQGNRPHLYE---GSGDVQKQETTNVAKKPDLEKT 971
 Db 1070 eedlerksidqepqlwstevaretvpeppespgslgsvdnretltslekesqelsl 1129
 Qy 972 MNW-----KDSVCFGQPRNDTN-----WQTPSSSYEQ----- 999
 Db 1130 gkwnvtrvedsqclqveegllqeeqheslrevkqelpssgnqgrwedvvgkavgqea 1189
 Qy 1000 -CATRQPHVLDIEFGMQEG-----LGYSWM---SISPRVDRVKNKNVPR 1041
 Db 1190 plattgvgtedkaelhrggggeeeaaagellqdivgeawslgssepkeqrvpaealdn 1249
 Qy 1042 RFRQGSVPREFRGOITPSPHELPGMLSGSSSAVQEHQDD--TOHQODEMKNKASHLQ 1100
 Db 1250 ---leggal-----evp---vaqsmpevterdedraqageqdsievtlgle 1289
 Qy 1101 --KTFDLNLSSECLRQSTKQNTIDGCLPRDRTAEDVVDPLSNSSNLIIVSNSS 1158
 Db 1290 aartglel-----eqevvgledprhfareaalpsalgsesvk----- 1326
 Qy 1159 NKEQTAVEYKETNATILREMKGTIADCKKTPSQWDSLRKQVEG-NEGROERKNMDSID 1217
 Db 1327 -----akiaqgleg---pdkp-----keagaldsgillelptsseale 1362
 Qy 1218 YEAIRRASISBAIKERGNMNLAVRIKDFLERIVKDHGGIDLEWLRSPPKAKD-- 1275
 Db 1363 cqghee---sesmgweeeas-----le--Lsdhegsdappr--ppteedeg 1405
 Qy 1276 YLLSIRGLGKSVPCVRLTLHLNLAFPVDNVGRIAVRMGVPLQPLPESLQHLHLEYP 1335
 Db 1406 aqaaltapgpkile-----pcspipiltah--elqp 1435
 Qy 1336 VLESIOKFLW 1345
 Db 1436 qaeqigqagw 1445

RESULT 5
 AAY85565
 ID AAY85565 standard; Protein; 2432 AA.
 XX
 AC AAY85565;
 XX
 DT 07-JUL-2000 (first entry)
 XX
 DE Human homologue of UNC-53 (HS-UNC-53/2) sequence.
 XX
 KW UNC-53; Caenorhabditis elegans; microtubule; neural regeneration;
 KW anticancer; anti-neurodegeneration; antifibrotic; anti-adhesive; human;
 KW antisclerotic; antimetastatic; anti-arthritis; autoimmune disease.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Region 1..89
 FT /note= "this region can be replaced with one of the
 FT three sequences shown in AAY85566 to AAY85568;
 FT this creates three variants at the N-terminus"
 FT
 FT Misc-difference 1018
 FT /label= Asp or Glu
 FT 1776..1778
 FT /note= "present or absent depending upon the allele
 FT from which the protein is translated"
 FT
 XX
 PN WO9963080-A1.
 XX
 PD 09-DEC-1999.
 XX
 PF 02-JUN-1999; 99WO-EP03848.
 XX
 PR 03-JUN-1998; 98GB-0011962.
 XX
 PA (JANC) JANSSEN PHARM NV.
 XX
 PI Luyten WHML, De Raemaeker MC, Geysen JJGH, Bogaert TAOB;
 PI Maerten LJS, Verhasselt P, Van De Craen M;
 XX
 DR WPI; 2000-116370/10.
 DR N-PSDB; AAA07836.
 XX
 PT Novel proteins and nucleic acids e.g. for treating neurodegeneration
 PS Claim 95; Fig 1d; 146pp; English.
 XX
 CC The invention provides vertebrate (human) protein homologue of a UNC-53
 CC protein of Caenorhabditis elegans. The UNC-53 binds to microtubules or
 CC their plus ends. The UNC-53 sequences are used to promote neural
 CC regeneration, revascularization and wound healing; also for treating
 CC neurodegenerative disease, acute traumatic injury, fibrotic disease and
 CC autoimmune diseases (e.g. rheumatoid arthritis and sclerosis). The UNC-53
 CC polynucleotides can be used for recombinant production of the proteins,
 CC as a source of probes for detecting allelic variants and polymorphisms,
 CC for sequencing genomic DNA and for detecting UNC-53 expression; and as
 CC source of therapeutic antisense sequences. Cells that express the
 CC protein are used to identify regulators of cell shape, growth, motility
 CC and migration. They can also be used to identify proteins that are
 CC involved in signal transduction pathways also involving UNC-53, and to
 CC identify compounds that alter attachment of UNC-53 to microtubules. A
 CC target gene coupled to a UNC-53 encoding sequence may be used to deliver
 CC the target gene to a cellular microtubule or its plus ends. The present
 CC sequence represents the amino acid sequence of the second human homologue
 CC of UNC-53, designated hs-UNC-53/2.
 XX
 SQ Sequence 2432 AA;

Query Match 2.3%; Score 208.5; DB 21; Length 2432;
 Best Local Similarity 18.7%; Pred. No. 1.8e-06;
 Matches 273; Conservative 171; Mismatches 547; Indels 469; Gaps 60;

CC The polypeptides are useful for providing analogues of gravin in the
 CC study of the modulation (e.g. blocking, inhibiting and stimulating) of
 CC interactions between gravin and kinase. The peptides are involved in the
 CC modulation of gravin-kinase interactions.
 XX
 SQ Sequence 1780 AA;

Query Match		2.1k;	Score 195;	DB 19;	Length 1780;
Best Local Similarity		17.8k;	Pred. No. 1.2e-05;		
Matches 332;		Conservative 262;	Mismatches 690;	Indels 584;	Gaps 84;
QY	105	ATKPNVGMNKSPEVTLKCRKALNFDLEPGDARQDSEISEIVQNS-----SGAN	158		
Db	23	aepepgggsaeapdt-----adpaa-asdpakllkngqlstingva	69		
QY	159	SFSIRDAIGTNGSFIDSVSDIKTNGLGAMNOPLE-----VSMGNQPD	203		
Db	70	eqdelsiqegdlng-----qkalgnggalnsqeeevivtevgqrdsedvserdsd	122		
QY	204	KLSTGAKLARDQOPDLLTRNQC--QPPVATON-----TQFPWENQQAWLQKN--OLIGFP	256		
Db	123	emactsavvhdttdgqeenrnieqipsesnlleltcq-ptesqandigfkvfkfvfk	181		
QY	257	FGNQPR-----MTIR-----NQQCLAMG-----NQQPMYLIGTPRPA	290		
Db	182	ftvkkktekptdvtqlitvkkgdegaagagdhqpslgagaesekspkq--stekpe	239		
QY	291	LVSNGOGLGGPOGNKRPIFNHOTCLPAGNOLVGSPTDMHOLVSTGCGQHGLLIKNQOP	350		
Db	240	etlkreq-----shaelsp-----paesggaavee--keegeeqkeqep	276		
QY	351	GSLIRGOOPCVPLIDQOPATPKGPTHLMQVATSMSSPGLRPHSOSOVPTTYLHVESVR	410		
Db	277	sk--saespsvptsetgstfkf--ftqwgagwrktsfrpkedeveas-----	333		
QY	411	ILNGTTCQSRAPAYDSLQDDIHQGNKYLSHEISNGNGCKKALPONSLPTPIMAKL	470		
Db	324	-----ekkkdeqekvteed--gkaevaseklcaseqahpqaesaheprisaey	373		
QY	471	EEAR-GSKROYHRAMGTETKHDNLNLAQOIAQSDQVERHNS-----TCVEYLDAAKKT	522		
Db	374	ekvelpseeqvgsgpseeekaplatevf-dekievhqeevvaevhvstveerteekt	432		
QY	523	KIQK-----VQENLHGNPPEVETIE-----DDPTDGARKGKNTASTISKA	563		
Db	433	eveetaagsvpaeevlgmadaepqaepakelviketcvsgedptqgadlspdekvisk-p	491		
QY	564	SKGNSPVKKTAEKECIVPKTPAK-----KGRAGRKSVPPPAHASEIOLWQPTTP	615		
Db	492	pegvsvsevmllssqermnkvggspklkiftstgklklsqkqkgrggdeesgehtqvp	551		
QY	616	KTPLSRSKPKGKRKSLQDSKARGPSGELLG-QDSIAELIYRMQNLYLGDKEREQFQNA	674		
Db	552	dspsdeeqkgessaspepe-----eltciekglaev-----qgdgeaeega	595		
QY	675	MVLYKGGG-----ALVPYES-----KKRPRP-KVDIDDETRINWLLMG-----	713		
Db	596	-----tsdgekkrregvtwafkkmvtpkkrvrrpseedkedeldkvksatlsstestase	651		
QY	714	-----KDEKEGDEKDKK-----EKWWE-----EER	736		
Db	652	mqeemkgsveepkpkrvkdvtsvsealivcgsskkrrrrrrssdeegpgkamgghq	711		
QY	737	RVFGRGRADSTIARHLVQGBRRRSPNKGSVVDSVIGVFLTQNVSDHLS--SAFMSLAARF	795		
Db	712	kadeagkdketgtdgfiagsehdpgggsspeqag---sptegegvstwesfrlvt--	766		
QY	796	PKLSSSREDERNVRSVV-----	813		
Db	767	prkksksleeksedsiaagsvehstpdtepgkeeswvskkflipgrkrkrpdkgkqeqap	826		
QY	814	VED--PRGCILNLNEIPSWQEKVQHPSPDMEVSGVDSGSKEQLR--DCSNSIERFNFLEKS	870		

Db	827	vedagptga-----nedds-----dvpavvplseydarekmeaaggqgkgaep--qka	875		
QY	871	IQNLEEVLSQDSFDPAPFQSCGRVSCSCSKSDAEFPTRCTETKTVSGTSQSVQGTSP	930		
Db	876	atevskelsesq-----vhmmaaavadgtraatifeerspswisavstpele	922		
QY	931	NLSDEICLOGNE-----RPHLYEGSGD--VQKQETTNVAQKKPD--	967		
Db	923	qveaeaaallteevlereleviaeeepptvtpeplpenreargdtvseaaeltpeavtaeta	982		
QY	968	-----LEKTMNWKDSVCFQGNRDNTNNQTTTPSSSYEQCATRQPHVLDIEDFGM	1015		
Db	983	plgseegeteasaaeeettemvsavsqiltdspdtteeatpvqeve-----ggvpdiee	1036		
QY	1016	QGEGLGYSWMSISPRVDRVKNKNVPRFRFGQGSVPREFTGQIIPSTPHPLPGMGLSGSS	1075		
Db	1037	rtqev-----lqavaevkvees-----qlpdtg--gpe	1062		
QY	1076	SAVQEHQDDTQHNOQDEMKNKASHLOK-----	1101		
Db	1063	dvlpqvq-raeaeerpeeaaasglkketdvvllkvdaqaekteftagkvvgqttesfek	1121		
QY	1102	--TFDLNLSSECLTROSST-----KQNTDCLPRDRTAEDVVDVPLSNNSLQNLV	1153		
Db	1122	apqvtesiesseelvtctcaetlagvksqemvmeqaiipd--svetpdtsetdstpvadfd	1180		
QY	1154	ESNSSNKEQTAVEYKETNATILREMKGTLADGKKPTSQWDSLRKDVEGNEGREGERNK--N	1211		
Db	1181	apdttkde-ivaiheenevhlvprgtaeavpaqkerppapssvfqeetkeqskmed	1239		
QY	1212	NMDSIDYE-AIRRASISEISEAIKERGMNMLAVRIKD--FLERIVKDHGIGD--LEWLR	1266		
Db	1240	tiehtkvevetvslsktegtqe--adqyadektdkvpffegl---egsidgtitvsr	1294		
QY	1267	ESPPDKAKDYLLSIRGLGKSVCEVR--LTLNLNAPPVDTNVGR-IAVRMGWVPLQPLP	1323		
Db	1295	ekvte-----valkgegteeeackddalelqshakspvverenvvqverekteap	1348		
QY	1324	ESLQLHLE--LYPVLESIQKFLWPLRCLDQRTLHYELHVOLITFGK-----	1368		
Db	1349	thvneeklehetavtveevskql-----lqtnvnpildgakevsslegspccc	1397		
QY	1369	-----VFCTKSRNCNACPMRGECRHFASAYASARLALPAPEERSL-TSATIPVPPE	1422		
Db	1398	lgqeeavctkiq-----vqseasftltaaaeeekvlgetaniletgetle	1443		
QY	1423	PVAIPMIEPLPLEKSLASCAPSNRENCEPIIEEPASPGQEC-----TEITESIEDI	1478		
Db	1444	pagahlv-----lee---ksseknedfaahpgedavtgpdcqakstpvivsattk	1495		
QY	1479	EDPDEIPTIKLNTIEQFGMTLREHMERMELQEGDMSKALVALHPTT--TSIPTPKLKNIS	1536		
Db	1496	sdlegkttslkws-----devdeqacvkvsvaledlepengillektskvlq	1549		
QY	1537	RL-----RTEHQVYEL-----PDSHRL-LDGMND-----KREDDP-----	1565		
Db	1550	niigtavdqvrteetatemltsetlqtqahvkdqsdagqetekegeepqasagdetpl	1609		
QY	1566	-----SPYLLAIWTGCTANSQA-----PPQKCGGKA	1593		
Db	1610	tsakeesestavqgahsdiskmseasektmtvevegstvndqgleevvlpseeegggag	1669		
QY	1594	SGKMCDFE 1601			
Db	1670	tksvpedd 1677			

RESULT 7
 AAB15380
 ID AAB15380 standard; Protein; 1780 AA.
 XX
 AC AAB15380;

Db 1020 qltdevthtqskvqqteeyl-----emkmbddlfeayirnkseadlllremen----l 1070
Qy 572 KKTAKKEKCIIVKTPAKGRAGRKKSVPPPAHASFIQLWQTPPKTPLSRSKPKGKGRKS 631
Db 1071 kgtmesvevkiadtkeleetirdkeqllhekkfyfqqamqtfipitplsdsippsk---- 1126
Qy 632 IQDSGKARGPSGELLCDQSIAGIIRVMONLYLGDKER-----EONAMVLYKGDGAL 684
Db 1127 -----lvgegn-----qdpie--indyhnlialaternnmvcleternsl----- 1165
Qy 685 VPYESKKRKPVPKVIDDITRIWMLMGKDEKGEDEKDKKEKWEERVRFRGRAD 744
Db 1166 -----keqvindntqlsl-----qagsieksdlqpkqglegev----- 1201
Qy 745 SFIARHLVQGRDRPSPKGVSVDSVIGVFLTQNVSDHLSSAFMSLAARPPPKLSSSRE 804
Db 1202 klllemell-----kghltdsqsliekqlenlevte-----klqtlqe 1240
Qy 805 DERNV-----RSVVVDDPGCILNLEIPSWQEKVQHPSPDMEVS 843
Db 1241 emknitiernelqtnfedlkaehdskldlsenleqstetqdelraaqeelreqkql--- 1297
Qy 844 GVDGSKQLRDCNSGIERENFLFKSITONLEEEVLSSODSFDPAIFOSGGRVGS CSCSK 903
Db 1298 -vds-frqqlldcs--vgisspn--hdavangeksvlgvnsllqsemllgerdelqtscka 1352
Qy 904 SDAEPTTTCETKTVSGTSQSVQTSQSPNLSDICLQGNR-----PHLYEGSGDVQKQ 956
Db 1353 lvselellrahvksvegenleittkklnglekeilgkseevlksmlenlkednknkq 1412
Qy 957 -----ETTNVAQKKPD-----LKTWNKDSVCFG--QPRN----- 985
Db 1413 aeeysskenqfsleevfsgsklvdeievkkaqlkaeerleikdrdyfelvqtantnlv 1472
Qy 986 DTNMQTTSSSYEQ-CATROPHVLIDIEFGMGEGGLGYSWMSISPRVDRVKNKNVPRFF 1044
Db 1473 egkietqladheedsidrrseemkivlgelkerngyllrigeekleisnkleilqke 1532
Qy 1045 RQGSVPREFTGQIIPSTPHELPGMLSGSSSAVOEHODDTQ---HNOQDQMNKASHL-- 1099
Db 1533 metsvllkddlqqklesllse--nllkenidttiklshsdtaqqlktqgqelqaknlai 1590
Qy 1100 -----QKTFDLNLSSE--CLTQSSSTKQNTDGLPRDRTAEDV--VDPLSN 1144
Db 1591 aasdnclptqeketsadcvhpleekillllteelhqtkneqekillhekeleaqvvelkce 1650
Qy 1145 NSSQLNILEVSNSS-----NKEQTAVEYKETAFTILREMGKTLADGKKPTSDWDSL 1195
Db 1651 vehlmskiesksleslqhekhdtqglialkqmqvvtqekkelqgthentlaevdhl 1710
Qy 1196 RKDVE-----GNEGRQERN-----NNMDSIDYEAIR-RASITSEISEAIKERM----- 1238
Db 1711 kenieIglnfneaqgkktkeqcllnnenkeleqshrlqceleelmkslkdkesaletlk 1770
Qy 1239 -NNMLAVRIKDFELRIVKDHGIDLEWLES-----PPDKAKDYLLSIRGLGKLSVCV 1291
Db 1771 eseqkvlnlndememvm-----lemeelknsgrtviaerdqlqddl-----resve-- 1816
Qy 1292 RLLTLHLNLAFFVDNMGVRIARMGVVLPQLPESIQHLHLELVPVLESIOKFLNPLRCKL 1351
Db 1817 -----msiet---qddllrkaqelqgqkdkvq-----eltsqisvlqe-----kisl 1856
Qy 1352 DQRTLYELHYQLITFGKVFCTKSRPNCNACPMRGECRHFASAYASARLALPAPEERSLTS 1411
Db 1857 enqml-----ynvatvketlserrddlnqskhlfsietlslslekefaleqaekdkada 1912
Qy 1412 A--TIPVPPESFPVAPIMELPLPLEKSLASGAPSNRENCEPIIEEPASQGCETEITE 1469
Db 1913 arktidi--tekislnleeqllqatnlketl-----yeresliqkeqialntehretlk 1966
Qy 1470 S-DIEDAYNEDPPEIPIKLNIEQFGNLTREHMERNM-ELQEGDMSKALVALHPPTTISI 1527
Db 1967 skdialgkmederdeaankvialtekmslleeqinenvtllkegegek----- 2014

Qy 1528 PTPKUKNISRLRTERHQVYVELDPDSHRLLD 1555
Db 2015 etfyqlrpskqgssqmeelreslktkd 2042

RESULT 9

AAR41043
ID AAR41043 standard; protein; 1786 AA.

XX AAR41043;
AC AAR41043;
DT 22-MAR-1994 (first entry)
XX CD4-EBAL75 fusion protein.
KW Merozoite; Erythrocyte Binding Antigen 175; malaria; HIV; env;
KW human immunodeficiency virus; envelope glycoprotein; hybrid protein;
KW red blood cell; erythrocyte; AIDS; molecular machine.

OS Chimeric Homo sapiens.
OS Chimeric Plasmodium falciparum.

XX Key Location/Qualifiers
FT Region 1..371
FT /note= "residues 1-371 of CD4"
FT Region 372..1786
XX /note= "residues 20-1435 of EBA-175"

PN WO9318160-A.

PD 16-SEP-1993.

PF 10-MAR-1993; 93WO-GB00505.

PR 11-MAR-1992; 92GB-0005276.

PR 08-JUL-1992; 92GB-0014481.

PR 24-JUL-1992; 92GB-0015829.

PR 16-SEP-1992; 92GB-0019562.

PR 03-MAR-1993; 93GB-0004311.

PA (PREN/) PRENDERGAST K F.

PI Prendergast KF;

DR WPI; 1993-303474/38.

XX Anti-viral fusion peptide(s) - comprise viral-binding component
XX and malaria merozoite red cell binding component, for treating
XX e.g. HIV, and hepatitis

XX Claim 9; Page 44-47; 69pp; English.

XX The hybrid protein NH2-CD4(1-371)-EBA175(20-1435)-COOH is a
XX specifically claimed example of a fusion protein of the invention;
XX it comprises at least part of the CD4 molecule fused to a peptide
XX from a malarial parasite merozoite protein with affinity for red
XX blood cells. The fusion protein can bind free HIV in the blood to
XX red blood cells and consequently reduce viral titre, prevent
XX transmission of the virus and improve safety of blood transfusions.

XX Sequence 1786 AA;

Query Match 2.18; Score 191.5; DB 14; Length 1786;

Best Local Similarity 18.9%; Pred. No. 2.4e-05;

Matches 180; Conservative 141; Mismatches 340; Indels 289; Gaps 42;

Qy 441 ILSEISNGNCKKALPSSSLPTPIMAK-----LEARGSKRQYHRAMGQTEKHDLNL 494

Db 651 mlsehkninnckn-ipqeelqitqwikewhgefllernrskpskcknt----- 702

Qy 495 AQQIAQSQDVERHNSSTCVELD-----AAKTKTKQVQVQENLHGMPPEVIEIE 543

```

703 -----lyeacekeidpcmkrydwlirskfwhltlskeyetqkvpknaeny---likis 754
544 DDPDGDARK-----GKNTASISKASGNSSPVKKTAKEKCIIVPKTPAK 588
755 enkndakvllnnncdaeyskycdkknttlivksvngndntik---ekrehldldfsk 811
589 KGRAGRKSVPPPAHAEIQIOWQPTPPKTPLSRKPKGKRSIQDSKARGPSGELLQ 648
812 fg--cdknsvd-----tnkvwecnpyi-----lstkd-----vc- 840
649 DSIAEIIYRMQNLVIGKEREQONAMVLYKGDGALVPYESK--KRPRPKVD-----ID 701
841 -----vpprrqelcgnidriyknllmikelaiaiaiesriilkyknkdkkevckii 895
702 DET-----TRVWN-----LLMGKDEKEGDEKDKKEK-----WWEERR----- 737
896 nktfadlrldiigdyndlsnrklivkintnskyvrhknkndkldfdevwkvjkkdwn 955
738 ----VFRGRADSFIAHMLVQGRDRFPWKGVSVDVSVIGVFLTQNVSDHLSsafmslAA 793
956 viswvfk---dktyckeddiqipqfrfwsewgddy----- 989
794 RFPKLSSSRDERNRSVVVE-----DPEGILNLNEIPSWQEKVQHPSDMEVSGVDS 847
990 -----cgdktmietlkyeckekpcedndckscnsykwiskkk----- 1029
848 GSKQLRDCNSGIERNFLEKSIQNLLEE-----EVLSSQDSFDPALF-----QSCGRVGS 898
1030 -----eeynkqakqyqyqgkgnnykmysefiksikevylkkysekcsnlnf 1075
899 CSCSK-----SDAEPPTTRC-ETKTV-----SGTSQSVQTSVGNLSDEICLOGNERPHLYE 948
1076 edefkeelhsdyknkctmcpekvdpisilrnnegtsqeaeppeenteila-hrtetpsise 1134
949 G-SGDVQKQ-----ETTNAQKKPDLEKTMNWKDSVCFQGP 984
1135 gpkneqkerddslskisvspensrpetdakdtsnlkikgvdism---pkavigssp 1191
985 NDTNWQTPSSSYEQCATRQPHVLIDIEDFGMGEGGLGYSMWSISPRVDRVKNKNVPRRFF 1044
1192 ndn-----invteqgdnisgvnsksplsdvrdpdkleled----- 1225
1045 RQGSVPREFTGQIIPSTPHELPGMGLSGSSAVQEHQDDTQHNOQDEMKNASHLQKTF 1104
1226 -qnseseetvnnhiskspinnngdsgsgsatvsessntglisiddrn-----gdtfv 1279
1105 ---DLLNSSECLTRQSTKONITDGC-LPRDRTAEDVDP-----LSNN-----SSLQNLV 1153
1280 rtqdtan-tedvirkenadkdedekadeerhstseslsspeekmltdneggnslnheev 1338
1154 ESNSNKEQTAVEYKETNATILREMKGTLDAGKKPTSQWDSLRKDVEGNEGRQERNKNM 1213
1339 kehtsndnvgqsggvmnmvekeikdtl---enpssld-----eg-kaheelsepn 1388
1214 DSIDYAIRRASISEISEAIKERGMNMLAVRIKFLERIVKHGGIDLE 1263
1389 ssdgdmsntpgplntseetterisnneykvneredertitkeyedivlk 1438
```

RESULT 10

AAP81187
ID AAP81187 standard; protein; 1284 AA.

XX AAP81187;

XX 04-DEC-1990 (first entry)

DT Sequence encoded by a gene fragment coding for major protein of an A-type
DE inclusion body and a promoter region in poxvirus.

XX Vaccinia virus; vaccine.

XX

Cowpox virus.
OS
XX Key Location/Qualifiers
FH Region 1..20
FT /note="Encoded by promoter region"
XX
PN EP261925-A.
PD 30-MAR-1988.
XX
PF 22-SEP-1987; 87EP-0308372.
XX 09-SEP-1987; 87JP-0223972.
PR 02-SEP-1986; 86JP-0222194.
XX (TOFU) TOA NENRYO KOGYO KK.
XX
XX Shida H, Funahashi S;
PI WPI; 1988-086185/13.
XX N-PSDB; AAN81538.
DR
XX Gene fragment coding A-type inclusion body in pox virus -
PT used in the construction of recombinant vaccinia viruses for use
PT as vaccines
XX
XX Disclosure; 3-10; 24pp; English.
PS
XX The gene is nonessential for proliferation of poxvirus and is homologous
CC with a corresp. gene of a vaccinia virus and can therefore be used for
CC construction of a recombinant vaccinia virus for use as vaccines. The
CC promoter present upstream of this gene is very strong and is adequate as
CC promoter for expression of an exogenous antigen cell.
XX
SQ Sequence 1284 AA;

Query Match 2.1%; Score 190; DB 9; Length 1284;
Best Local Similarity 18.4%; Pred. No. 1.8e-05;
Matches 223; Conservative 159; Mismatches 387; Indels 442; Gaps 56;

QY 469 KLEEARSKQYHRAMQTEKH-----DLNLAQOIAQSDVERHNSSTCTVEVLDAAKTKI 524
DB 414 kceecnggeykteleearkkitelneleiskirtler-----dsvykt-- 461
QY 525 QKVVOENLHGMPPEVIEDDPTDGARKKNTASISKAGNSSPVKKTAKEKCIIVPK 584
DB 462 -----eridrltkeikeld-----ingtdgdsdsseidkktirelsdrer----- 506
QY 585 TPAKKGRAKRKSVPPPAHASEIQIOWQPTPPKTPLSRKPKGKRSIQDSKARGPSGE 644
DB 507 -----emrteleere-----ldtird-gkvegs--- 527
QY 645 LLCQDSTAEIIRYMONVLYGDKEREQONAMVLYKGDGALVPYESKRRKPPKVDIDDET 704
DB 528 --cq-----relelrmwl-----kdrddldraeidkrr 554
QY 705 TRIWNLLMGKDEKEG---EEKDKKE-----KWEEERRVFRGRADSP-IARMH 751
DB 555 nvewelsrlrldikecdkyekidkaktisnyvristleseiakyqqdrdtlsvrire 614
QY 752 LVQGDRRFPWKGVSVDVSVIGVFLTQNVSDHLSsafmslAAARPPKLSRDERNRS 811
DB 615 leeertrvrdlesrlde-----ctrnqetqevdalrsirelenkladmesggg--- 665
QY 812 VVVEDPEGCILNLNEIPSWQEKVQHPSDMEVSGVDSKESQDLRDCNSGIERNFLEKSI 871
DB 666 -----nltelsrlgski---sdle-----rqlrecrgnate-isrlqyrl 701
QY 872 QNLEEEVLSSQDSFDPALFQSCGRVSGVSGVSGVSGVSGVSGVSGVSGVSGVSGVSG 931
DB 702 tdlterql-----ndcrrnne---nnadteremqrlrditdlerq----- 738

```
Qy 932 LSEDTCLQGNRPHLYEGSGDVQKQETTNVAOKKPDLEKTMNWKDSVCFGQPRNDTNWOT 991
Db 739 lsd--crrnn-----esadmer-emgrldrindldrqln----- 771
Qy 992 TPSSSYEQCATRQPHVLDIEDFGMGEGGLGYSWMSISPRVDKKNKVPRRFFROGGSVP 1051
Db 772 -----ec-----krdngtcs-----seevnrlktrf----- 792
Qy 1052 REFTQGIIPSTPHELPGMGLSGSSSAVOEHODDTQHOHQDQNMNKASHLQKTFDLNLSNE 1111
Db 793 -----rdlersleicskdeselaysksselgrar-----eqisnlq 828
Qy 1112 ECLTRQSTKQNTQICLPRDRTAEDVDVPLSNSSSLQNLIVESNSS----- 1158
Db 829 eslrreresdk--tdsyyrreltre-----rnkiveleleknkcfctdnhakyi 874
Qy 1159 ---NKEOTAVEYKETNATILRMKCTFLADGKKPTQOWDSLAKDV---EGNEG----- 1204
Db 875 deinskktrisdlerqlaacsngsgsdmdgdykreieslkrrelaecrrgnngshsdcky 934
Qy 1205 -----ROE--RNKNMNSIDYEAIRRASISEAIEKERGMNMLA--VRIKDFLERIVK 1255
Db 935 ydeearveevklrlqeltql-hedlkrateskndsykrelgrakvievekeleryfd 993
Qy 1256 D-----HGGIDLEWLRESPPDKAKDYLLSIR--GLGLSKSVECVRLTLHLNLAFFVDT 1305
Db 994 dsrlaackrhgd---emlr-----kiadleklrdgngngngctsscefer----- 1038
Qy 1306 NVGRTAVRMGWVPLQPLPESLQHLLELYPVLESIQKFL-WPRL--C--KLDO----- 1353
Db 1039 --krlav-----leevrkmetikslkfmefdlqdcadklidrekerim 1083
Qy 1354 RFLYELHYQLITFGKVFCTKSRPNCNACPMRGECRHFASAYASARLALPAPAE----- 1406
Db 1084 kaerdlerei-----arkncggnpcerelesersnvrkleyqldaekvfykr 1133
Qy 1407 -----RSLTSATIPVPPSPFPVPAIMIEPLPLEKSLASGAPSNRENCEPIIEE 1456
Db 1134 elerdrylssryltssdp-dekplpnytfpriev-----epittedtepkpve 1181
Qy 1457 PASPOQECTEITESDIEDAYNEDPDPTIKLNTIEQFCGWTLEHMERNMELQEGDMSKA 1516
Db 1182 vvpssdvtepiisgy-----tpsvdaepe-----hpdlseiyqtsvs 1218
Qy 1517 LVALHPTTSTIPTPKLNISRLRTEHQVYELPDSSR-----LLDGMKREPDDESP 1567
Db 1219 qvavtpp-----pkpetqiseyqdyselysasnntesknvsefslayldldkl--ddide 1272
Qy 1568 YLLAIWTPTGET 1578
Db 1273 yllnnlmpekt 1283
```

RESULT 11
AAB12454
ID AAB12454 standard; Protein; 2453 AA.

AA AAB12454;
AC
XX
DT 24-OCT-2000 (first entry)
XX
DE HNRCR protein sequence.
XX
KW Human; HNRCR; nuclear receptor coreceptor.
XX
OS Unidentified.
XX
PN CN1250094-A.
XX
PD 12-APR-2000.
XX
PF 06-OCT-1998; 98CN-0120919.

```
PR 06-OCT-1998; 98CN-0120919.  
XX (XINH-) XINHUANGPU FUDAN GENE ENG CO LTD SHANGHA.  
XX  
XX Yu L, Tu Q, Zhao Y;  
XX WPI; 2000-400830/35.  
DR N-PSDB; AAA60630.  
XX  
XX Preparation of new human keron acceptor co-repressor coding series and  
XX the polypeptide  
XX  
XX Example 2; Fig 2; 58pp; Chinese.  
PS  
XX  
XX The present invention describes a human homologue of nuclear receptor  
XX coreceptor (HNRCR). The present sequence represents an HNRCR protein  
XX sequence used in comparison with the human HNRCR.  
XX  
XX Sequence 2453 AA;  
SQ  
Query Match 2.1%; Score 189; DB 21; Length 2453;  
Best Local Similarity 17.8%; Pred. No. 6.4e-05;  
Matches 287; Conservative 204; Mismatches 553; Indels 568; Gaps 67;  
Qy 2 QSIMDSSAVNATEAQND-----GSRQDVLEFDL-----NKTPOQKPSKRX- 43  
Db 735 esapspsveaakssedssenaasrgntepvaeleattdpacaaspssavpttkpaeres 794  
Qy 44 -----RKMPKVY-----VEGPKRKKPKPAELPKVVEGPKRPRKRAATQE 86  
Db 795 veaqvtdsaetaepmdvdheecgaegssvldppaptkadsvdpe-----mqvpe 845  
Qy 87 KVKSETGSAKKNLKESATKPPANVGMNSKSPVTLKSKRKALNFLENPGDARQDS 146  
Db 846 ntaskgedakderldestekteardv-----vvaqierp-epgsdd 890  
Qy 147 ESEIVONSNGANSFSE-----IRDAIGTNGSFLDSVSQIDKTNGL-----GAM 190  
Db 891 ssatscdegvdgepergrvfpmdakpalltpggsil--isspikpnllldlpqlghraav 948  
Qy 191 NQPLEVSGMGNQDKLSTGAKLA-----RDQOPDLTLTNQOCC 227  
Db 949 lppm---vscpcnlpigtvpvsgyalyrhlkhamhesalleeqrqreqvdlcsrsts- 1004  
Qy 228 FVATQNTQFPMENQOAWLQMKNLIT-GFPFGNQOP-----RMTIRNQOP 271  
Db 1005 pcstks--pnrewevlqphqvtlpegrvrlpttrpppplpssktttvasekp 1061  
Qy 272 CLAMGNQPMYLIQTPRPALVSGNQOLGGPQGNKRPFLNHOTCLPAGNQLYGSPTDMHO 331  
Db 1062 sfimgggsisq---gtpgtylsshnday--pgeapkp-----svgsisiglprrqgs 1107  
Qy 332 L-----VMSGGOQHGLLTKNQOPGSLIRGQQPCVPPLIDQOPATPFGFTHL 377  
Db 1108 tkaapltyikgeefsrqsqpegllvraqheg-vvrgtagav----qegsitrg-tpa 1161  
Qy 378 NQVATSMSS-----PGLRPHSQSQVPTTYL----- 403  
Db 1162 skisvetisslrgsitqgtpalpqagiptealvkgpvsmpleesspekvreeaaskghv 1221  
Qy 404 -----HVESVSRILNGTTGTCQRSRAP-----AYDSLQODIHQ----- 437  
Db 1222 lyegkshgihlydniknaregt---rsprtahemslkrseavegsklqgmmsrespsv 1277  
Qy 438 -----NKYILSHEISNG-----NGCKKALPONSSILPPI 466  
Db 1278 apleglicralprgshsdlkervtvisgmsimggtprataesfedgk--ypkqlkrespp 1335  
Qy 467 MAKLEEARSGKRYH-----RAMQTEKHDLNLAAQIAQSODVERH-----N 508  
Db 1336 irafegaitkgydygittikemgrsihelpqrdi-ltqesrktpevvqgstprliegsls 1394
```

```
QY 509 SSTCEYLDAAKTKIQKVOENLHG---MPPEVIEIEDDPTDGARKG----- 554
D 1395 ggtplkfdnsggaikhnvkslitgpsklprgmleivpenikvvekyedvkagepvr 1454
QY 555 --NTASTISKG-----ASKGNSSP-----VKTAKEKCIYPKT--- 585
D 1455 arhtsvsvsgpsvirstheapkaqlspgldysarrtpvsyqntisrgspmmnrtsdv 1514
QY 586 PAKKGRAGRKSPPPAHASIQWLQWTPP-----PKTPLSRSPKPKGRKS----- 631
D 1515 sssksasherkstltptqresipakspvgvdpivshpsfphhrssaagevyrshlpth 1574
QY 632 ---IQDGGKARGSGELLCQDSIAEI-----IYRMQN---LYLGDKEREQONAMVL 677
D 1575 ldpampfhraldpaaayllqrlsptpgypsqyqlyamentrtqilndyitsqmqvnlr 1634
QY 678 YKGDGALVPYESKKRPPK-----VDID-----DETTRI----- 707
D 1635 pdvtgrispreqplglpypatrliditnmpptilvphaggtstppmdrityipgtqtf 1694
QY 708 -----WNLLMGKDEKDEKDKKKKWEERRVFRGRADSFIAARMHL 752
D 1695 pprpynaaslsphgpthlaaasaererererererereriaaapadlyl 1754
QY 753 VQG-----DRFSPMKSGSVVDSVI----- 771
D 1755 rpgseqqrpqshgvyrspsvrtqetllqrpqsvfgtngtstvitpdpqlrimpl 1814
QY 772 ---GVFLTONVSDHLSSAFMSLAARPPPKLSSSRED-----ERNVRS--- 811
D 1815 psqgplsqqipasyryntaadaalaaydaasapqmdvsktkeahaaarleenlrssa 1874
QY 812 VVDEPE-----GCILNLNEIPSWQEKVQHPDMEVSGVDSGS-----KEQ 852
D 1875 avseqqlqeknlvekrsvctvctsaalpsgkqphasvvyseagkdkgppkrsyree 1934
QY 853 LRDCSNGIERFNFLEKSIQNLVEELSSQDSFDPFQSCGRVSGSCSCSDAEFFPTR 912
D 1935 lrtgkttitaanfivli---trqiasdkda-----rergsgsdssslsshr 1981
QY 913 CETVTSCTSOVOTGSPNLSD-----EICLOGNERPHLYEGS-GDVOKQ 956
D 1982 yet-----asdaievispaasppqekpqayqdpdmvkanqaenestryegpllhvrsq 2036
QY 957 ETTNVAOKKPD-----EKTWNWKDSVC-----FGOPR-----NDTNWQ 990
D 2037 qespsqgqpllpsssegmgqvrthrlitladhicqitqdfarnqvpssqgstsfq 2096
QY 991 TTPSS-SYEQCATROPHVLDIEDFGMOGEGLYSWMSPISPRV-DRVKNKV--PRRFRQ 1046
D 2097 tpsalsstpvrtktsryape-----sqstvlhprpgrvpenlvdkrsgrp 2147
QY 1047 GGSVPRFETGOIIPSTPHELPGMGLSGSSSAVOHQDDTOHQODENKASHLOKTFDL 1106
D 2148 gkspersh---ipseye-----pisppqgpavhekqds-----lll 2182
QY 1107 LNSSECLTROSKQONITDGLPRDRTAEDVVDPLSNSSLONLVSNSS 1158
D 2183 sqryvdaeqrsdrspgsisylpsfftkiestspmv-kakkgqefrkluss 2233

RESULT 12
AA70232
ID AAR70232 standard; Protein; 1435 AA.
XX
AC AAR70232;
XX
DT 22-SEP-1995 (first entry)
DE
DE P. falciparum SABP.
XX
KW SABP; stalic acid binding protein; binding domain; merozoite;
KW malaria; therapy; vaccine.
```

```
XX Plasmodium falciparum.
OS
XX
FH Key Location/Qualifiers
FT Domain 1..616
FT /label= Binding_domain
XX
XX WO9507353-A.
XX
XX 16-MAR-1995.
XX
XX 07-SEP-1994; 94WO-US10230.
XX
XX 10-SEP-1993; 93US-O119677.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
XX Wellens TE;
XX
XX WPI; 1995-123427/16.
XX
XX N-PSDB; AAQ83525.
XX
XX New erythrocyte binding domain polypeptide(s) - isolated from
XX Plasmodium binding proteins, used in diagnosis, treatment and
XX prevention of malaria
XX
XX Disclosure; Page 39-41; 81pp; English.
XX
XX Sequences from the SABP gene (given in AAQ83525) were PCR amplified,
XX expressed on the surface of COS cells and tested for erythrocyte
XX binding to identify the binding domain polypeptide. A prefd. SABP
XX binding domain comprises residues 1 to about 616 of the SABP protein
XX (AAR70232). Recombinant binding domain was expressed in E. coli,
XX yeast, mammalian, insect, and in vaccinia virus and adenovirus-
XX infected cells. It provides protection against P. falciparum.
XX
XX Sequence 1435 AA;

Query Match 2.1%; Score 188.5; DB 16; Length 1435;
Best Local Similarity 18.3%; Pred. No. 2.8e-05;
Matches 185; Conservative 144; Mismatches 345; Indels 337; Gaps 42;

QY 441 ILSEIHSNGCKKALPQNSSLPTPIAK-----LEARGSKRQYHRAMQTEKHDNLN 494
D 299 mlsehknfmimcnkn-ipqeelqtqwikewhgeflldhrsklpkskknnt----- 350
QY 495 AQQTAAQSDQVERHNSSTCVELD-----AAKTKIQKVOENLHGMPPEVIEIE 543
D 351 -----lyeacekecidpmkyrdwiirskfewhtlskeyetqkvpkenaeany---likis 402
QY 544 DDPTDGARK-----GKNTASISGASKGNSSPVKKTAKEKCIYPKTPAK 588
D 403 enkdakvsi1lnncdaesyksyckdhttlvksvlnngndntik---ekrehidlddfsk 459
QY 589 KGRAGRKSPPPAHASEIOLWOPTTPKTPLSRSKPKGKGRKSIQDSGKARGPSGELLQ 648
D 460 fg-cdksndv-----cntkvwcknpyi-----lstkd-----vc- 488
QY 649 DSIAEIIYRMONLYLGDKEREQONAMVLYKGDGALVPYESK---KRKPRPKVD-----ID 701
D 489 -----vprrrqelcplgnidrydknllmikehilataiesrilkrkyknkddkevckii 543
QY 702 DET-----TRWN-----LLMGKDEKDEKDEKKEK-----WVEERR----- 737
D 544 nktfadiirdiigtgyndlnsrklvgkintnskyvhrnkndklfrdewkwvkkdawn 603
QY 738 ----VPRGRADSFIAARMHLVQGDRRFSPWKGSVVVDVIGVFLTONVSDHLSSSAFMSLAA 793
D 604 viswvfk---dktvckeddienipqffrfsewgddy----- 637
QY 794 RFPKPLSSSREDERNVRSVVVE-----DPEGCILNLNEIPSWQEKVQHPDMEVSGVDS 847
```

```
Db 638 -----cqdktkmletikveckepecdndcksknsykewiskkk----- 677
Qy 848 GSKEQLRDCSNGIERFNFLEKSIQNLEE-----EVLSSQDSFDPAIF-----QSCGRVGS 898
Db 678 -----eeynkqakqyqeyqkgnnykmysefksikpevyllkkyseksninf 723
Qy 899 CSCSK-----SDAEFTTRC-ETKTV-----SGTSOSVOTGSPNLSDEICLOQNERPHLYE 948
Db 724 edefkeelhsdyknkctmcpvkdvpsilrnnedtsqeaapeenteia-hrtetpsise 782
Qy 949 G-SGDVQKQ-----ETTNVAQKKPDLEKTMNWKDVSVCFGQPR 984
Db 793 gpkgneqkerddslskisvspensrpetdaktdsnllklkgdvdisn---pkavigasp 839
Qy 985 NDTNMQTTPSSSYEQACATR-----QPHVLDIEDFGM-QGEGGLYSWMSISPRV----- 1031
Db 840 ndninvteggdnisgvnskpsldvdpdkleedqnsdeseetvvnhiskspsinnngds 899
Qy 1032 -----DRVKNKNVPRFRFGQGSVPRE 1053
Db 900 ggsatvsesssntglslddrngdtfvrtqtdtantedvirkenadkdedekgadeerh 959
Qy 1054 FTGOIIPSTPHELPGMLSGSS---SAVOEHQDDTQHNOQD-----EMNKASHLQKTF--- 1103
Db 960 ssesisspeekmldcneeggnslnheevkehtsntsdnvqggqgvinmvekeikdtlenp 1019
Qy 1104 ---LDLLSSSESLTRQSTQKQNIITDGLPRDRTAEDVVDPLSNSSSLQNILVESNSNK 1160
Db 1020 sssldegkaheelsepnlsdqdmntpgpldntseetterisnneykvneredertitk 1079
Qy 1161 EQTAVEYKETNATILREM-KGTLDGKKPTSQWDSLRKDVE-----GNEGRO----- 1206
Db 1080 evedivlk---shmrresaddgelydensdlstvdnsedesadaekmkgndtsemshssghi 1136
Qy 1207 --ERNKNMNDSD-----YEAIRASISEISEATKER 1236
Db 1137 esdqkndmktvgdltgthvqneisvptgeideklreskesklnhaeere 1187

RESULT 13
AAW22477
ID AAW22477 standard; Protein; 1435 AA.
XX AC AAW22477;
XX DT 11-SEP-1997 (first entry)
XX DE Silaia acid binding protein.
XX KW DBL gene family; SAPP; sialic acid binding protein; merozoite; malaria;
XX KW Duffy antigen binding protein; DABP; erythrocyte; var-1; var-2; var-3;
XX KW var-7; vaccine; therapy; immune response; Plasmodium.
XX OS Plasmodium falciparum.
XX PN W09640766-A2.
XX PD 19-DEC-1996.
XX PF 07-JUN-1996; 96W0-US09508.
XX PR 07-JUN-1995; 95US-0487826.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
XX PI Wellens TE;
XX WPI; 1997-052231/05.
XX DR N-PSDB; AAT72888.
XX PT New malaria vaccines - contains cysteine-rich DBL family protein
```

```
PT binding domains homologous domains of the Duffy and sialic acid
XX binding proteins
XX Example 1; Page 37-40; 96pp; English.
CC This sequence represents the full length sialic acid binding protein
CC (SAPP). SAPP and the Duffy antigen binding protein (DABP) are soluble
CC proteins that appear in the culture supernatant after infected
CC erythrocytes release merozoites. DABP and SAPP mediate the binding of
CC merozoites and schizonts to the erythrocyte surface. These proteins are
CC necessary for erythrocyte invasion by the parasite. This sequence can be
CC used in the compositions of the invention. The compositions are for the
CC treatment and prevention of malaria, and comprise either a nucleotide
CC sequence or encoded polypeptide of the var-1, var-2, var-3 or var-7 genes
CC of the DBL gene family, a family of genes having homology with conserved
CC regions of DABP and SAPP. The compositions are used for the treatment and
CC prevention of malaria. They are also used in the preparation of vaccines
CC for inducing a protective immune response in a mammal to Plasmodium
CC merozoites (especially Plasmodium falciparum or Plasmodium vivax).
```

XX Sequence 1435 AA;

Query Match 2.1%; Score 188.5; DB 18; Length 1435;
Best Local Similarity 18.3%; Pred. No. 2.8e-05;
Matches 185; Conservative 144; Mismatches 345; Indels 337; Gaps 42;

```
Qy 441 ILSHSISNGCKKALPONSSLTPTIMAK-----LEARGSKROYHRAMGOTEKHDLML 494
Db 299 mlsekhninnckn-ipqeelqitqwikewhgefillerdnrsklpkscknnt----- 350
Qy 495 AQOIAQSDVERHNSSTCVEYLD-----AAKTKIOKVQVENLHGMPPEVIEIE 543
Db 351 -----lyeacekecidpcmkrydwiirskfwhetlsketyetdkvpknaeny---likis 402
Qy 544 DDPDTGARK-----GKNTASISKASKGNSPVTAKTEKICVPTPAK 588
Db 403 enkdakvslllnnccdaeyskycdckhttlvksvlnngndntik---ekrehlidddfsk 459
Qy 589 KGRAGRKXSVPPPAHASEIQLMQTPPTPKTPLSRSPKPGKGRKSIQDSGKARGPSGELLCO 648
Db 460 fg--cdknsvd-----tntkvwecknpyl-----lstkd-----vc- 488
Qy 649 DSIAELIYRMQNLVLDKEREQEQNANWLYKGDGALVPYESK---KKKPRPKVD-----ID 701
Db 489 -----vpprrgeicignidrydknllmikahilalaiesrllkrkyknkdkkevckli 543
Qy 702 DET-----TRWN-----LLMGKGEKEGEDEKDKKEK-----WDEERR----- 737
Db 544 nktfadiirdiigtdywndlnsrkivgkintaskyvrnkknkndklfrdewkvikkdwn 603
Qy 738 -----VFRGRADSFIAARMHLVQGRDRFSPWKSGVSDVIGVFLTONVSDHLSSAFMSLAA 793
Db 604 viswvfk---dktvckeddienipqfrfwsewgdy----- 637
Qy 794 RPPPKLSSREDERNVRSVVVE-----DPEGCIINLNEIPSWQEKVQHPDMEVSGVDVS 847
Db 638 -----cqdktkmletikveckepecdndcksknsykewiskkk----- 677
Qy 848 GSKEQLRDCSNGIERFNFLEKSIQNLEE-----EVLSSQDSFDPAIF-----QSCGRVGS 898
Db 678 -----eeynkqakqyqeyqkgnnykmysefksikpevyllkkyseksninf 723
Qy 899 CSCSK-----SDAEFTTRC-ETKTV-----SGTSOSVOTGSPNLSDEICLOQNERPHLYE 948
Db 724 edefkeelhsdyknkctmcpvkdvpsilrnnedtsqeaapeenteia-hrtetpsise 782
Qy 949 G-SGDVQKQ-----ETTNVAQKKPDLEKTMNWKDVSVCFGQPR 984
Db 783 gpkgneqkerddslskisvspensrpetdaktdsnllklkgdvdisn---pkavigasp 839
Qy 985 NDTNMQTTPSSSYEQACATR-----QPHVLDIEDFGM-QGEGGLYSWMSISPRV----- 1031
Db 1137 esdqkndmktvgdltgthvqneisvptgeideklreskesklnhaeere 1187
```

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 12:42:17 ; Search time 20.61 Seconds
(without alignments)
1689.960 Million cell updates/sec

Title: PCT-US01-13059-2
Perfect score: 9089
Sequence: 1 MQSIMSSAVNATEEQND.....PRPLMARLHPPASKLNKNT 1729

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 2014635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	210.5	2.3	1805	1 US-07-853-913-2	Sequence 2, Appli
2	195	2.1	1780	1 US-08-769-309A-5	Sequence 5, Appli
3	195	2.1	1780	3 US-08-994-570-5	Sequence 5, Appli
4	188.5	2.1	1435	2 US-08-568-459A-4	Sequence 4, Appli
5	188.5	2.1	1435	2 US-08-487-826B-4	Sequence 4, Appli
6	187.5	2.1	1346	2 US-08-635-121-2	Sequence 2, Appli
7	185.5	2.0	2409	6 5180808-2	Patent No. 5180808
8	185	2.0	2441	3 US-08-961-739-2	Sequence 2, Appli
9	183	2.0	2441	1 US-08-194-468-2	Sequence 2, Appli
10	182	2.0	1863	1 US-08-598-591-2	Sequence 2, Appli
11	182	2.0	1863	1 US-08-798-691-2	Sequence 2, Appli
12	182	2.0	1863	1 US-08-798-691-6	Sequence 6, Appli
13	182	2.0	1863	3 US-08-825-487A-2	Sequence 2, Appli
14	182	2.0	1863	3 US-08-825-487A-6	Sequence 6, Appli
15	182	2.0	1863	4 US-09-074-476-2	Sequence 2, Appli
16	182	2.0	1863	4 US-09-074-476-4	Sequence 4, Appli
17	179	2.0	1663	5 PCT-US93-07261-16	Sequence 16, Appli
18	177.5	2.0	3248	1 US-08-353-700-1	Sequence 1, Appli
19	177.5	2.0	3248	5 PCT-US95-16216-1	Sequence 1, Appli
20	177	1.9	1863	2 US-08-603-753D-2	Sequence 2, Appli
21	177	1.9	1863	4 US-09-099-753-2	Sequence 2, Appli
22	177	1.9	1863	4 US-08-986-106-2	Sequence 2, Appli
23	175	1.9	1210	1 US-08-320-559-26	Sequence 26, Appli
24	175	1.9	1210	3 US-08-545-860D-26	Sequence 26, Appli
25	175	1.9	1210	5 PCT-US94-04496-26	Sequence 26, Appli
26	174	1.9	1588	5 PCT-US93-07261-11	Sequence 11, Appli
27	173	1.9	2414	1 US-08-227-536-2	Sequence 2, Appli

28	173	1.9	2414	5 PCT-US95-04682-2	Sequence 2, Appli
29	172.5	1.9	1187	1 US-08-320-559-28	Sequence 28, Appli
30	172.5	1.9	1187	3 US-08-545-860D-28	Sequence 28, Appli
31	172.5	1.9	1187	5 PCT-US94-04496-28	Sequence 28, Appli
32	172.5	1.9	2842	1 US-07-741-940-7	Sequence 7, Appli
33	172.5	1.9	2842	1 US-08-289-548A-7	Sequence 7, Appli
34	172.5	1.9	2842	1 US-08-452-654-7	Sequence 7, Appli
35	172.5	1.9	2843	1 US-08-452-655B-2	Sequence 2, Appli
36	172.5	1.9	2843	1 US-08-452-655B-7	Sequence 7, Appli
37	172.5	1.9	2843	4 US-08-450-582-2	Sequence 2, Appli
38	172.5	1.9	2843	4 US-08-450-582-7	Sequence 7, Appli
39	172.5	1.9	2973	2 US-08-821-355A-7	Sequence 7, Appli
40	172.5	1.9	2973	2 US-09-003-687A-7	Sequence 7, Appli
41	172.5	1.9	2973	4 US-09-136-605-7	Sequence 7, Appli
42	172	1.9	1852	1 US-08-425-061-24	Sequence 24, Appli
43	172	1.9	1852	2 US-08-825-886-24	Sequence 24, Appli
44	172	1.9	1863	1 US-08-425-061-16	Sequence 16, Appli
45	172	1.9	1863	1 US-08-480-784-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-07-853-913-2
; Sequence 2, Application US/07853913
; Patent No. 5338839
; GENERAL INFORMATION:
; APPLICANT: McKay, Ronald D.G.
; APPLICANT: Lendahl, Urban
; TITLE OF INVENTION: Nestin Expression As An Indicator of
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/853,913
; FILING DATE: 19920319
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/660,412
; FILING DATE: 22-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/603,803
; FILING DATE: 25-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/201,762
; FILING DATE: 02-JUN-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/180,548
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-4641AAAA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1805 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-07-853-913-2

Query Match 2.3%; Score 210.5; DB 1; Length 1805;

Best Local Similarity 18.7%; Pred. No. 3.2e-08;

Matches 283; Conservative 216; Mismatches 564; Indels 447; Gaps 72;

QY	60	PRKPAELPWVVEGRPKPRKPAATAQEKVSK--ETGSAKKKNLKESATKPPANVGDMSN	111
DB	159	PRRP-----PAPHRIPGPAPEYEDLARELGEVWRGAVRDYQERVAHMESSLQGARE	210
QY	118	KSPVET--LKSCRKALNFLEPGDARGDSESEIVONSSGANSSEIRDAI--GTGNSGF	174
DB	211	RLSQAVRGARECR-----LEVQQLADRDLSLOERREALQRLGCRW	251
QY	175	LDSVSDIKTN--GLGAMNOPLEVSMGNOPDKLSTGAKLARDQOOPDLLTRNOOCOPPVATQ	233
DB	252	QDRLOATDRFQLAVEALEKEGQGLSQIAQILLEGQQLAH-----LKMSLSELEVATY	303
QY	234	NTQFPMENOQAWLOMKNOLLIGFPFGNOOPRMTIRNOQOCLAMGNOPMYLICTPPPALVS	293
DB	304	RTLLEAENSR--LQTPGRGSQSLGFLDPKL-----KNPF--LGIPEDQYL--GSVULPALP	354
QY	294	GNQOLGGPOGNKRPI--FLNHOTCLPAGNOLYGSPTDMHQLVMSTGGQOHLIKNOOPG	351
DB	355	TSFPSPLPNTLETPTVTAFLKTQEFELQARTPLAS--TPIPPISEAP-----CPPN	402
QY	352	SLIRGOOPCVPLIDQOPATP---KGFTHLNQWVAIAMSSPG-----LR	391
DB	403	AEYRAQEVPLSLIQTQAPEPLMKATVPSSSAILPELEPGGKQOGHPFDDTLTSLATNLN	462
QY	392	PHSOSQVPTTYLHVESVSRILNCTGTGTCORSRAPAYDSLQODIHOGNKYI-----LSHEI	446
DB	463	PHH-----PT-----LEAKGSESSESRVSSI--FOEDGQIWEIWEKEADIEVKV	505
QY	447	SNNGCKKALPONSSLPPTIMAKLEBARGS-----KROYHRAMGQTEKHDNLNAQOIAQSQ	502
DB	506	ENSSAQK-----TQESGLDT-----BETQDSQGLQKETLTKALG--EPLMSLUKIQNYETA	554
QY	503	DYERHNSSP-----CVEYILDAAKTKIKQKVQENL-----HGMF--PEVIEIEDDDTD	548
DB	555	GRNCNSSTEGHGLTLEGPEKEQIPLKSLEBKNVSEKTELVNGVPVIELLGGKEDTRTE	614
QY	549	-----GARK-----KNTASISKAGKGN-----SSPVAKT	574
DB	615	DQELMSPKTLKRESSLGRQSEVVWRPSKEGNEKWSHTAFKESQHPGLFPGAEQDQMLERL	674
QY	575	AEKEKCIIVKTPAKG-----RAGRKSVPPPAH-----ASETLQWQTPPK	616
DB	675	VEKQDSQFRSPPEEDQEAECRPLQENQBPGLYEBAEQOILERLIEKESQESLSRPEED	734
QY	617	TPLSRKPKGX-----GRKSIDQSGKARGPSGELLQCDSIAETIYRMONLYLGDK--ERQOE	671
DB	735	QEAGRSLOKENQBPGLYEBAEQMLER-----LIBEKESQLSKSPERNQTKGKPLERNQ	789
QY	672	QNAWLYKGDGALVPYESKKRPRKPVIDDDTTTRINLLMGKGDEKBDG--EEKDKKKEK	730
DB	790	KSRLYLEENQOETFPVLESRNQPLRSLEVEEBEQIRVAPLEKVSQDSGLSGLAEENVQPLR	849
QY	731	WHEEE-----RRVFRGRADSFIAARMHLVGDRFRSPWKGVSDSVIGVFL-----TONVS	780
DB	850	YLEEDDCINKSLEEDKTHKSIGSLERNQDSIILP-----QESETQVSLRPPPEEDQRIV	904
QY	781	DHLSSAFMSLAAREFPPKPLSSSREDERNV--RSVVVEDPEGCILMLNIPSWQEKVQHPQS	838
DB	905	NHLEKES-----QEFSRSSEEEQVMERSLGENHE-----SLSSVEKEDQWVE--S	949
QY	839	DNEVSGVDSG-----SKEQLRDCNSGIERFNFLEKSIQNL-----EE	876
DB	950	QLEKESQDSGKSLDESQETFGPLBKENAESLRSIAGQDQBEQKLEQETQOTLRAVNGEQ	1009
QY	877	EVLSSQDSDFP-----AIFQSGCRGVGSCS--CSKSDAEFFPTR--CETKTVSGTS--	922

: INFORMATION FOR SEO ID NO: 5:

; SEQUENCE CHARACTERISTICS:

SEQUENCE CHARACTERISTICS:
LENGTH: 1780 amino acids

```

, LENGTH: 1780 amino acids
:
: TYPE: amino acid

```

TYPE: amino acid
TOPOLOGY: linear

TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-769-309A-5

Query Match	2.18:	Score 195:	DB 1:	Length 1780:
-------------	-------	------------	-------	--------------

Query Match: 2.1%, Score 193, DB 1, Length 1780,
Best Local Similarity: 17.8%, Pred. No. 6.5e-07:

2536	Local Similarity	17.0%	Freq: NO: 0.56 077,
Matches	332: Conservative	262: Mismatches	690: Indels
		584: Gaps	84: Gaps

QV 105 ATKKPANVGDMNSKSP E V T I . K S C R K A I . N E D I . E N P G D A B O G D S E S E T V O N S ----- S G A N 158

QY IUJ AIKRPANVGDMSNRSEFVILKSCRAALNFULENPGDARQGDUSESEIVQNS-----SGAN 138

DB 23 AEPEPSGGPSAEAAPDTT-----ADPAIA-ASDPATKLLQKNQSLTINGVA 69

150 CRYPTOPOLYGRAPHIC CIPHERS

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

Db 70 EQDELSLQEGDLNG-----QKGALNGQGALNSQEEEEVIVTEVGQRDSEDSERDSK 122

QY 204 KLSTGAKLARDQQPDLTRNQCCQFPVATN-----TQFPMENQAWLQMKN--QLIGFP 256

[illegible]

NOT VTC A TTY TO CONTACT THE DIRECTOR OF THE FBI FOR INFORMATION ON THE
FBI'S POLICY ON THE USE OF FORCE BY LAW ENFORCEMENT OFFICERS.

Qv 257 FGNOOPR-----MTIR-----NOOPCLAMG-----NOOPMYLIGTPRPA 290

Db 182 FTVVKKDKTEKPPDTVOIITVKKDEGECAACAGDHODPSICAGEAASKFSERKO--STFKDE 239

DB 182 FTVKRDKTEKPTVQLLTVKRDEGEAGAGDHQDPSLGAGEAAKSEEPKQ--STERPE 239

QY 291 LVSGNQQLGGPQGKRPFLNHQTCLPAGNQLYGSPTDMHQLVMSTGGQQHGLLIKNQQP 350

[illegible]

QY 351 GSLIRGQQPCVPLIDQQPATPKGFTHLNQMVATSMSSPGLRPHSQSQVPTTYLHVESVS 410

Db 277 SK--SAESPTSPVTSETGSTFKKF--FTQGWAGWRKKTSTFRKPKEDVEAS----- 323

.....

[illegible]

Db 324 -----EKKKEQEPEKVDTEED---GKAEVASEKLTASEQAHPOEPAESAHEPRLSAEY 373

OV 471 EEAR-GSKROYHRAMGOTEKHDLNLAOOTAOSODVERHNSS-----TCVEVI.DAAKKT 522

QY 471 EAK-GSKKQIHKAMGVIEKHDLNLAQVIAQSQDVERHNSS-----TCVEYLDAAKRT 322

DB 3/4 EKVELPSEEQSGSGPSEKPAPLATEVF-DEKIEVHQEEWAEVHVSTVEERTEEQKT 432

07 523 K10K --- WWOENI HCMRBEVIEF --- DDDTDCADKVCUNWASTCZCA 562

[illegible]

Db 433 EVEETAGSVPAEELVGMDAEPQBAEPAKELVKLKETCVSGEDPTQGADLSPDEKVLSP-P 491

Qy 564 SKGNSSPVKKTAEKEKCVPKTPAK-----KGRAGRKKSVPPPAHASEIQLWQPTPP 615

—

—

 **

 **

 **

—

—

..

—

..

..

..

..

—

—

..

(

.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

8f UIC KIF ESAYGRTKOROKNCTIQDSSRKAROFSGEETC QDSIARELTITIKMQNLILBOPKEREQEUNA 074
| | : : | | : : | : | : | :
: | : : | | : : | : | : | :

RESULT 3

US-08-994 - 570-5

US-08-994-370-3
: sequence 5. Application IIS/08994570

Sequence 3, Applicant
Patent No 5090929

GENERAL INFORMATION:
APPLICANT: Scott, John D.,
APPLICANT: Nauert, Brian J.,
APPLICANT: Klauck, Theresa M.,
TITLE OF INVENTION: Protein Binding Domains of Gravin
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower/233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/994,570
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: No. 609029and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/33451
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1780 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-994-570-5

Query Match 2.18; Score 195; DB 3; Length 1780;

Best Local Similarity 17.88; Pred. No. 6.5e-07; Matches 332; Conservative 262; Mismatches 690; Indels 584; Gaps 84;

QY 105 ATKPANVGMNSKSPVETLKSCRKALNFDLENPCDARGDSESEIVQNS-----SCAN 158
DB 23 AEPEPGGSAEAPDT-----ADPAIA-ASDPATKLLQKNGQLSTINGVA 69
QY 159 SFSEIRDAIGTNGSFLDSQIDKTNGLGMNOPLE-----VSMGNQPD 203
DB 70 BDELSLQEGDLG-----QKALNGQALNSQEEVIVTEVGQDSDVSESDSK 122
QY 204 KLSTGAKLARDQPDLLTRNOOC-QFPVATON-----TOFPMENOQAWLOMKN--OLIGFP 256
DB 123 EMATKSAVVDITDDGQENRNIQIFSSSNLELTQ-PTESQANDIGKPKVFKFGFK 181
QY 257 FGNOQPR-----MTIR-----NOQPCLAMG-----NOQPMYLIGTRPA 290
DB 182 FTVKDKTEKPDVTOLLTVKDKDEGAAGADHDPSLGAGEAASKESEPKQ--STEKPE 239
QY 291 LVSGNOQLGGPGNKRPLFLNHQTLCPAGNOLYSPDMDHOLYMTSGQOHLIKNQOP 350
DB 240 ETLKREQ-----SHAISIP-----PAESGOAVEEC--KEEGEEKQKEP 276
QY 351 GSLIRGOQPCVPLIDQOPATPKGFTHLNQMVATSMSSPGLRPHFSQSQVPTTYLHVESVR 410
DB 277 SK--SAESPTSPVISEGTFKFKF--FTQWAGWRKKTSTFKPKDEVEAS----- 323
QY 411 ILNQTGTQCRAPAYDSLOQDTHQGNKYILSHSINGNGCKKALPONSLSPTPIMAKL 470
DB 324 -----EKKKEQPEKVDTEED---GKAEVASEKLTASEQAHPQEPAESAEHPRLSAEY 373
QY 471 EEAR-GSKROYHRAWGTEKHDLNLAQOIAOSQDVERHNS-----TCVEYLDAAKT 522
DB 123

DB 374 EKVELPSEEQVSGQSPSEKPAPLATEVF-DEKIEVHQEEVVAEVAHVSTVEERTEQKT 432
QY 523 KIQK-----VVOENLHGMPEVIEIE-----DDPTDGARKGKNTASISKGA 563
DB 433 EVEETAGSVPAEELVGMDAEPQEAEPKELVKLKETCVSGEDPTQGDLSDEKVLSP-K 491
QY 564 SKGNSPVKKTAEKEKCIVPKTPAK-----KGRAGRKSVPPPAHAHAIQLWQTPPP 615
DB 492 PEGVSEVEMLSQERMKVQSPKLLFTSTGLKLLSGKKQKGRGGDESGHTQVPA 551
QY 616 KTLPSRSKPKGKRKSIQDSGKARGPSGELLC-QDSIAEIIYRMNLYLGDKEREQONA 674
DB 552 DSPDSQEQEQGESSASSPEEPE-----EITCLEKGLAEV-----QODGEAEEGA 595
QY 675 MVLKGDG-----ALVPYES-----KKRPRP-KVIDIDETTRINLLMG----- 713
DB 596 ---TSDGEKREGVTWASFKNMVTPKKRRPSESDEKDELKVKRSATLSTESTASE 651
QY 714 ---KGDEKGEDEKDKK---EKWE-----EER 736
DB 652 MQEEMKGSVEEPKPEEPKRVDTSVSWEALICVSGSKRRRRSSDEEGPKAMGGDHQ 711
QY 737 RVFRGRADSFIAHMLVQGRFRSPWKSVVDSVIGVFLTONVSDHLSS-SAFMSLAARF 795
DB 712 KADEAGDKKETGTGILAGSQEHDPGGSSPEQAG---SPTGEGYSTWESFKRLVT-- 766
QY 796 PPKLSSSREDERNVRSVV-----813
DB 767 PRKKSKELEKSEDSIAGSGVEHSTPTDEPGKEESVSIKKFTIPGRKKRPDKQKQOAP 826
QY 814 VED---PEGCIIILNLEIPSWQEKVQHPSDMEVSGVDSKQELR-DCSNSGIERNFLKS 870
DB 827 VEDAGPTGA---NEDDS---DVPVAVPLSEYDAVEREKMEQAQQAQKAEQPE--QKA 875
QY 871 IONLEEEVLSQDSFDPALFOSCGRVGSCSKSDAEFFTRCTETKTVSGTSQSVQFGSP 930
DB 876 ATEVSKELSESQ-----VHMAAAVADGTRAATIIERSPSWISASVTEPLE 922
QY 931 NLSDEICLOGNE-----RPHLYEGSD--VOKQETTNAQKKPD--- 967
DB 923 QVEAEAAALLTEEVLEREVIABEPPVTPELPENREARGDTVVSEALTEAVTAATAG 982
QY 968 -----LEKTNMKDVSFCFGPRNDTNMQTTTPSSSYEQCATRPHVLDIEDFGM 1015
DB 983 PLGSEECTEASAEETEMVSQVLSQDTPDTEATPVQVEE-----GVDPDIEQER 1036
QY 1016 QGEGCLGYSMWISIPRVDRVKNKNVPRFRFGGSGVPREFTGQIIPSTPHELPGMGLSGSS 1075
DB 1037 RTQEV-----LOAAEKVKEES-----QLPGTG--GPE 1062
QY 1076 SAVQEHODDTQHNQODENMKASHLOK-----1101
DB 1063 DVLOPVO-RAEAERPEQAEASGLKKTEDVVLKVDAQEAKEPTQGVKVVQOTTPESFEK 1121
QY 1102 --TFDLNLSSECLTRQSTT-----KQNTDGLPRDRAEDVDVDPVLSNNSLQNILV 1153
DB 1122 APQVTESSIESSELVTCQAEETLAGVKSQEMVMEQAIIPD-SVETPTDSETDGTSPVADF 1180
QY 1154 ESNSSNKEQTAVEKKTENATILREMGKTLADGKKPTSDWSLKRKDVBNEGROERNK--N 1211
DB 1181 APGTTQDE-IVEIHEENEVHLVPVRGTEAEAVPAQERPPAFSSFFVQETKEQSKMED 1239
QY 1212 NMSIDYE-AIRRASISEISEAIKERMNNMLAVRIKD--FLERIVKDHGID--LEWLR 1266
DB 1240 TLEHTDKESVEIVTSILSKTEGTQE--ADQVADEKTKDVPFEGL---EGSIDGTIVSR 1294
QY 1267 ESPDPKADKYLISIRGLGKSVCEVR--LITLHNLAPPVDTNVGR-IAYRWGVPLOPLP 1323
DB 1295 EKYTE-----VALKGEETEEACKDDALEQSHAKSPSPVEREMVVOVEREKTEAEP 1348
QY 1324 ESLQLHLLLE--LYPVLESIQKFLWPRCLKDQRTLYELHVQLITFGK----- 1368
DB 1349 THVNEEKLHETAVTVSEEVSKQJ-----LOTVNVPIIDGAKEVSSLEGSPPPC 1397

Db 1137 ESDQKNDMKTVDLGTTHVQNEISVPVTGEIDKLESKESIKHAEER 1187

RESULT 5

US-08-487-826B-4
; Sequence 4, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellem, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1435 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-08-487-826B-4

Query Match 2.1%; Score 188.5; DB 2; Length 1435;
Best Local Similarity 18.3%; Pred. No. 1.6e-06;
Matches 185; Conservative 144; Mismatches 345; Indels 337; Gaps 42;
Qy 441 ILSHEISNGCKKALPQNSLPPIMAK-----LEEARGSKQYHRAMGQTEKHDNL 494
Db 299 MLSEKNNINCKN-IPQEEQLITQWKEWHGPELLERDNRSLPKSKCKNNT----- 350
Qy 495 AQQTAQSDVHRHNSSTCVYLD-----AAKTKIQKVQVONLHGMPPVEIE 543
Db 351 -----LYEACEKECIDCKMYRWIIRSEFWHTLSYETQKVPKNAENY---LIKIS 402
Qy 544 DDPTDGARK-----GKTAISKASKGNSPVKTKAEKCIVPKTPAK 588
Db 403 ENKNDKAVSLLNNCDAYESKYCDCKHTTTLVSVLNGNDNTIK---EKREHIDLDDFSK 459
Qy 589 KGRAGRKSVPPPAHASEIOLWQPTPPKTPLSRSPKPKGKRSIQDSKARGPSGELLQ 648
Db 460 FG--CDKNSVD-----TNTKWECKNPI-----LSTKD-----VC- 498

Qy 649 DSAETIIRQNLVLDKEREQEQNAMVLYKGDALVPYESK--KRPRPKVD-----ID 701
Db 489 -----VPPRQELCLGNIDRIYDKNLLMIKEHILAIAYESRIILKRYKKNKDEVCII 543
Qy 702 DET-----TRIWN-----LLMGKGEKEGEDEKDKKEK-----WWEERR----- 737
Db 544 NKTADIRDIIGDGYWDLNLRKLVGKINTNSKYVHRNKNKDLFRDEWMVKIKDVWN 603
Qy 738 ----VERGRADSEIARMHLVGGDRFSPWKGVSVDVIGVFLTONVSDHLLSSSAFSLAA 793
Db 604 VISWVFK---DKTVCKEDDIENIPQFRFSEWGGDY----- 637
Qy 794 RFPKLSRREDERNRVSVVVE-----DPEGCILNLEIPSWQEKVQHPFSDMEVSGVDS 847
Db 638 -----CQDKTKMIETLAVECKEPCEDDCKSKNSYKWKSKK----- 677
Qy 848 GSKQLRDCNSGIERNFLEKSTONLEE-----EVLSSQDSFDPALF-----QSCGRVGS 898
Db 678 -----EYNKQAKQYQYKGNVKNMYEFKSIKPEVYLKRYSEKCSNLF 723
Qy 899 CSCSK----SDAEPTTRC-ETKT-----SGTSOSVOTGSPNLSDEICLQGNRPHLYE 948
Db 724 EDEFKEELHSDYKNKCTMCPEVDVPIIIRNNEQTQEAPEPTEIHA-HRTETPIS 782
Qy 949 G-SGDVQKQ-----ETNTVAQKKPDLEKTMNMKDSYCFQGP 984
Db 783 GPKNEQKERDDDSLSKISVSPENSREPETDAKDTSNLLKLGVDVISM--PRAVIGSSP 839
Qy 985 NDNWQTPSSSYEQCATR-----QPHVLDIEDFGM-QGEGLYSWMSSISPRV----- 1031
Db 840 NDNINVTQGDNIGVNSKPLSDVDRPKKELEDONSDESEETVNVNHSKSPSINNGD 899
Qy 1032 -----DRVKNKNVPRFRFGGSGVPRE 1053
Db 900 GSGSATVSESSSNTGLSIDDRNGDTFVRTQDTANTEDVIRKENADKDEKGADEH 959
Qy 1054 FTQIIPSTPHELPGMGLSGSS---SAVEHQDDTQHNQD---EMNKASHLQKTF--- 1103
Db 960 STSELSPEEKMLTDNEGNSLNHEEVKEHTSNSDNVQSGGIVNMNVEKELKDTLENP 1019
Qy 1104 --LDLLNSSECLTROSSTKQNTDCGLPRDRTAEDVDVPLSNSSLOLILYESNSNK 1160
Db 1020 SSSLDEGAHEELSEPNLSDDQMSNTPPGLDNTSEETTERISNNEYKVNEDERTLT 1079
Qy 1161 EQTAVEYKETNATILREM-KGTLDAGKKPTSQWDSLKQVE---GNEGRO----- 1206
Db 1080 EYEDIVLK---SHMNRSDGELYDENSIDLSTVNDSEDEDAEAKMGNDTSEMHSN 1136
Qy 1207 --ERNKNMDSID-----YEAIRRAISEISEIAKER 1236
Db 1137 ESDQKNDMKTVDLGTTHVQNEISVPVTGEIDKLESKESIKHAEER 1187

RESULT 6

US-08-635-121-2
; Sequence 2, Application US/08635121
; Patent No. 5910442
; GENERAL INFORMATION:
; APPLICANT: Gelman, Irwin H.
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10112-0228
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/635,121
;; FILING DATE: 19-APRIL-1996
;; CLASSIFICATION: 514
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Clark, Richard S
;; REGISTRATION NUMBER: 26,154
;; REFERENCE/DOCKET NUMBER: A30558 - 165/33603
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 212-408-2558
;; TELEFAX: 212-765-2519
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1346 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; HYPOTHETICAL: NO
;; ANTI-SENSE: NO
;; FRAGMENT TYPE: internal
;; ORIGINAL SOURCE:
;; US-08-635-121-2

Query Match 2.1%; Score 187.5; DB 2; Length 1346;
Best Local Similarity 17.5%; Pred. No. 1.7e-06;
Matches 278; Conservative 192; Mismatches 513; Indels 607; Gaps 70;

QY 353 LIRGQPCV-----PLIDQAPATPKGFTHLMQVATS-----MSS 387
Db 63 LKMSREMCVSGDHTQLDLSPEETLPHKPGIYSEVEMLSQERIKVQGSPLKLFSS 122
QY 388 PGLRPHS-----QSQPTTYLHVESVSRIINGTTGTCQSRAPAYDSLOODTHQ 436
Db 123 SGLKSLGKKQKRGKGDEEPGYOIHITES-----PESADEQK 163
QY 437 GNKYILSHEISNGCKKALPONSSLPIMAKL--EEARGSKROYHRAMGQTEKHDLNL 494
Db 164 GESSASSPEETTCLEGP-----LEAPMGKLLKELLRGEKKR-----KDHSLGI 211
QY 495 AQIIAQSDVERHNSSTCVLEYLDAAKTKIQVVOENLHGMPEVIEEDD-----545
Db 212 LQKDGDTQETVRRPS-----ESDKEELEKVKSATLSSTDSTVSEMDQEVKTVGEQK 264
QY 546 PTDGARKGNTAS-----TSKGASK-----GNSSPVKKTAKEKCK-----580
Db 265 PEPKRRYDTSVSWPALICVSGSSKKRARKASSSDIRGPRTLGGQSGQSRGGQQRQSRRTD 324
QY 581 IVPKTPAKKGRAGKKSVPV---PAHASEIQLMQP-----TPPKTPLSRSPKGGKRSKI 632
Db 325 AVPASTQEQDQOGSSPEPAGSPSEGGVSTWSEFKLVTPRKKSKEKEAGRTLV 384
QY 633 QDSG-----KARGPSGELLCCODSTAEIIRYMONLYLGDK-EREQEQNAM 675
Db 385 VGAGCPLRSNRVEKNLGFPLRNSPDG-----GRKGQMGREQAT 424
QY 676 VLYKGDGALVPYESKKRPRKVDIDETTRIWNLLMGKGEKEGDEKDKKKKEWBE 735
Db 425 VEDSG-----PVEINEDEP-----DVPVAVPL-----SEYDAVEREKME-----458
QY 736 RVFRGRADSFIAHMLVYQDRRFPWKSVDVIGVFLTQNVSDHL--SSAPFMSLAA 793
Db 459 -----AQGNALPSCVGVVSEELSKTLVHTVSVAVIDGTRAVTSVEE 501
QY 794 RFPPLKSSS-----RDERNRSVVVED-----PEGCIILNLNIPSWQ 831
Db 502 RSPSWISASVTEPLEHTAGEAMPPEVTEKDIIAEETPVLQTLPPEG-----549

RESULT 7

5180808-2

; Patent No. 5180808

; APPLICANT: RUOSLAHTI, ERKKI I.

QY 832 EKQVHPSDMEVSGVD-----SGSKE-----OLRDCSNS 859
Db 550 -KDAH--DMVTSEVDFTSEAVTATSETSEALRTEEYVTEAGAEETDMVSVAVSLDSTPT 607
QY 860 GIER--FNFLEKSIQNLLEEVLSSQDSFPALFQSCRGVSGSCSKSDAEFFTRCETKT 917
Db 608 TEETATPVQVEGGVLDTEEEERQTQ-----ALQAV-----ADKVEESOVATQTVQRT 657
QY 918 VSGTSQSVOTGSPNLSDEICLOGNERPHLYEGSDVQVQOETTNVAQKPDLEKTMNWKDS 977
Db 658 GSKALEKVE-----EVEEDSEVLASEKEKDMPKGPVQEAAGAEHLAQ-----699
QY 978 VCFGQPRNDTNWQTPSS-----SYEOCATRQPHVLDIEDFGMOGEGGLGYSNMSISP 1029
Db 700 -----GSETQATPESLEVPETADVHDVATCQ--VIKLOQLMEQ-----AVAP 741
QY 1030 RVDVRKNKNVPRFRFGSGSVPREFTGQIIPSTPHELPCMGLSGSSSAVQAEHQDDTHNQ 1089
Db 742 ESSETLIDS-----ETNGSTP-----LADSDTADGTOQDETIDSQ 776
QY 1090 QDEMKNASHLOKTFDLDLNSSEECITROSSTKONTIDGCLPRDRTAB-----DVEDPLSN 1145
Db 777 DSKATAAVRQSOV-----TEEAATAQKEPSTLPNNVPAQEEHGEHPGRDVLPTQOE 830
QY 1146 SSLQNLIVESNNSKEQTAVYEKETNATILREMKGTLDGKK-----PTSQ-- 1191
Db 831 LAAAAVPVWQKTEVQGEVGDW-----LDGEKVKEEQEVFVHSGPNSSOKA 875
QY 1192 ----WDLRKDVEGNEGROERNKNMNDSDIDEA-----IRRASISESEAIKEKGMNM 1241
Db 876 ADVTYDS-----EVMGAGQOEKESTEVQSLSEEGEMETDVEKEKRETRPEQVSEGEQET 932
QY 1242 LAVRIKDFLERIVKHGGIDLEWLRESPPDKAKDYLLSIRGLGLKSVCEVRLTLHLNL-- 1299
Db 933 AA-----PEHERNYGKPVLTLDMPSS--RGKALGSLG 964
QY 1300 --AFPVDNVGRIAVRMGWVPLQPLPESLQHLLELYPVLESIQKFLPRLCKLQORTLY 1357
Db 965 SPSPDQDKAGCIEQV-----QSLDTTVTQ--TAAEAVEKI-----999
QY 1358 ELHYQLITFGKVFCKSPNCNACPMRGECHRFASAYASARLALPAPERSLTSA-----1412
Db 1000 ----ETVWISE---TGESPEC-----VGAHL-LPA-EKSSATGGHWTIQ 1034
QY 1413 ----TIPPPESPFPVAPMIELPLLEKSL-----ASCAPSNRENCEPIIEEPASP-- 1460
Db 1035 HAEDIVPLGPES-QAESIPILVTPAP-ESTLHPDLQGEISASORSESEEDKPDAGPAD 1092
QY 1461 GOECTEI-----TESDIEDAYNEDPDEIPTIKLNIQFGMT-----1497
Db 1093 GKESTAIKVLKAEPEILELESKNKIVLNIQTAVDQFARTETAPETHAYDSOTQVPAM 1152
QY 1498 -----LRHEMERNMELQEDMSKALVALHPTTISI-----1527
Db 1153 RLDSREPNRCHTKMKVAKMHPVPOPREDLOVLTVLEAWLSSEMLAALAVESAGVKVST 1212
QY 1528 ---PTPKLKNISRLRTEH-----QVYELPDSHRLDGMKREPDDPSPYLLAINTPGETA 1579
Db 1213 KLPPQPKDQK-----EHAADGQPLQSLAQAEAVSNLTKESPDINGPKL-----1256
QY 1580 NSAQPEQKCGGK--ASGKMPDETCSE-----CNLSRE-----ANSOTVRGT 1620
Db 1257 -----TEERCPOKLSRKKKCLPSQSKRTRPQKRKTRCSORETQWQNKMLVAHCTSVRP- 1310
QY 1621 LLIPCRTAMRGSFPLNGTYFQVNFELFADHE 1650
Db 1311 ---ECENKSONKMLLLGPWTKISEPMSRE 1337

; CURRENT FILING DATE: 1997-10-31
; EARLIER APPLICATION NUMBER: US 194,468
; EARLIER FILING DATE: 1994-02-10
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2441
; TYPE: PRT
; ORGANISM: Mus
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)-(2441)
; OTHER INFORMATION: xaa = Any Amino Acid
US-08-961-739-2

Query Match 2.0%; Score 185; DB 3; Length 2441;
Best Local Similarity 17.5%; Pred No. 7.9e-06;
Matches 342; Conservative 222; Mismatches 685; Indels 708; Gaps 84;
QY 5 MDSSAVNATEATEQNDGSRQDYLEFLDKNT-----PQQPKSRKRKFMKPVVVEGPKRK 59
Db 287 MGATGVNPOLASKQSMVNSLPAPFTDIKNTSVTVPMNSQLQTSVGIPTQAIATGPTAD 346
QY 60 P--RPAELPKVVVEGPKRKPKKAATQEKVSKETGSAKKNLKESAT-----KKPANVG 113
Db 347 PEKRLIQOQLVLLHAHKCORREANGEVXACSLPHCRMTKNVLMHMTHCQAPKACQVA 406
QY 114 DMSNKSPEVT-LKSCRK---ALNFDLENPGDAROGD---SESEIVONSSGANSFSFSEIRD 165
Db 407 HCASSRQIISHKNCNTRHDCVCLPLKNASDRNQQOITILGSPASQIONTIGS-----V 459
QY 166 AIGGTNGSFLDSVSQIDKTN-----GLGAMNQPLEY----- 196
Db 460 GAGQQNATSLSNPNPIDPSSMGRAYAAALGLPMNQPTQLQPVPGQQPAQPPAHQOMRT 519
QY 197 --SMGNQPKLSTGAKLARDQPDLLTRNQOCQFPVATONTQPPHENQAWLQMKNOLLG 254
Db 520 LNALGNMSPVPAAG--ITTDQOQPNLISEALSPLTSLGATN-----PLMND-----G 564
QY 255 PFGNQOQPMTRNOQPCLAMGNQOPMY-----LIGTPRAL----- 291
Db 565 SNSNGISLSLTIPTAAPSPSTGVKRGWHEHVTDLRLSHLVHLVQLVQAIPTTDPALKDRR 624
QY 292 -----MENLAYAKKVEGDMYESANSRDEYHLLAEKIYKIQKELEEKRRTRLHKQILGNOPAL 684
QY 300 GPQGNKRPIFLNHTCLPAGNOLYGSPTDMHOLVMSSTGGQHQHLLIKNOOPGSLIRGQOP 359
Db 685 PASGAQPPVIPPQAQVRPPNGPL---PLPVNRMQVSQG-----MNSFNPMSLGNVQLP 734
QY 360 CVPLIDQOQPATPKGTHLNQMV-ATSMSS-PG-LRPHSQSQOVPITYLHVESVRLNGT 415
Db 735 -----QAPMGFRAASPMNHSVQMSMASVPGMAISPSRMPQPPN-----MMGT 777
QY 416 TGTGORSRAPYDSIQDDITHQGNKYLILSHSINGNGCKKALPONSLSLPTPIIMAKLEEARG 475
Db 778 HANNIMAQAQTONQE-----LPQN----- 796
QY 476 SKROYHRAMQTEKHDLNLAQOIAQSDQVERHNSSTCVE-----YLDAAKTKIKQVQE 530
Db 797 ---QFPSSSGAMSVNSVGMGQPAAGVSGQGEPAALPNLMLAPOLPCPPVPTOS 853
QY 531 NLHGMPPEVIEEDDP-----TDGARKGKNTASISKASK 565
Db 854 PLHPTPPASTAAGMPSLQHTAPGMTPPQAPATQPTSPVSSGQTPTTPTGCVPSAAOT 913
QY 566 GNSSPVKTAKEKICVPTPAKGRAGRKSVPPPAAHASEIQLWQPT-----PPKTPLS 620
Db 914 QSTPTVQAAAAQAVTPQOPTVQ-----PPSVATPQSSSQOQPTPVHTQPGCTPLS 963
QY 621 RSK-----PKGKRKSIQDSGKARGPSGELLQCQDSIAEIIYRMQNLVYLGDKEREQON 673

Db 964 QAAASIDNRVPTPTSTVTSAAETSSQPGDPVPM-----EMKEVQT-----DAAEPE----- 1010
QY 674 AMVLYKGDGALVPYESKRRKPRKPKVDIDDETTRINWLLMGKDEKDEKDKKKKWE 733
Db 1011 -----PTES-KGEPRSEMEED-----LAGSSQVKEETDTTEKSPMEV 1049
QY 734 EERRVFRGRADSFIAARMHLVQDGRFSPWKGVSVDSVIGVFLTQNVSDHLSSAFMSLAA 793
Db 1050 EEK-----PEVKEAKEEENSNDTASQSTSP 1078
QY 794 RFPPLKSSSREDER-----NVRSVVVEDPEGIL-----NLNEIPSWQEKVQHPSDMEV 842
Db 1079 SQPRKKIKFPEELROALMPTLEALYRQDPESLPFRQVDPQLLGPDYDIYKPNPMDLST 1138
QY 843 --SGVDSGSKEQ-----LRDCSNSGIERFNELEKSIONLEBEVSSQDS 884
Db 1139 IFRKLDTGQYQFPWQYVDDVRLMFNNANWLNKTSRV--YKCSKLAEFQEIE----- 1190
QY 885 FDPALFQSGRVSQSCSKSDAEFPTRTCTKTVSTQSOTGSPNLSDETC-----L 938
Db 1191 -DP-VMQS---LGVCGRKYEFSPQTLCCYKQLCTIPRDAAYVSYQNYHFCCGKCFTEI 1245
QY 939 QGNERPHLYEGSDGVQKQETTNNVQAKKPDLEKTMN----- 973
Db 1246 QG-ENVTL---GDDPSQPQTIS---KQDFEKKNKNDTLDPEPFVDCCKEGRKMHIQICVLH 1298
QY 974 ---WKDS-VC-----FCQPRNDTNW---QTPPSSSY-----EQCATRQPHVLIDIED 1012
Db 1299 YDIWPSGFVCDNCKKTGCRPRKENKFSAKRLQTFGLGHLEDRVKNKFLRRQNHPEAGEV 1358
QY 1013 F-----GMOGEGLYSWMISIPRVDRVKNKNVPRR-----FFRQGSVPR 1052
Db 1359 FVRVWASSDKTVEVKPMKSRFVDSGEMS-----ESFPYRTKALFAFEIDGVDV 1408
QY 1053 ERTGQIIPST---PHELPGMG-----LSGSSSAVOR--- 1080
Db 1409 CFFGMHVQDTALIAPHQIQGCVYISYLDIHFPRPRCLRTAVYHEILIGYLEVKKLVV 1468
QY 1081 -----HODD---TOHQOODEMNKASHLQ-----KTFDLNLSSEECTRQSS 1119
Db 1469 TAHIWACPPSEGDDYIFHCHPPDQKIPKRLQEWYKMLDKAFARIINDYKIFKQAN 1528
QY 1120 TKONITDGLCPDRDRAEDVVDPLSNSSQNLILVESNSSNKEQTAVEYKETTATILREMK 1179
Db 1529 -----EDRLTSAKELPYFEGDFWPNVLEISKELEQEEERKEEETAASSETPE 1577
QY 1180 GTLADGKPKTSQWDSLKDOVEGNEGRQERNKNMDSIDYEAIRRASISSEISAKEP--- 1236
Db 1578 GSQGSK-----NAKKNNKTKNKK-----SSISRANKKKPSM 1611
QY 1237 -GMNNMLAVRIKDFLER-----IVKDHGGIDLEWLRESPPDKAKDYLSIRGLGLKSVE 1289
Db 1612 PNVSDNLSQKLYATWEKKHKEVFEVHLAG---PVISTQPPVDPDPPLS----- 1658
QY 1290 CVRLLTLNLAPVDNTVGR-----IARVMGWVLPQPLPESLQLHLELYPVLESIQKFL 1344
Db 1659 -----CDLMDGRDAFLTLARDKH-----EFSSLRRSK 1686
QY 1345 WPRLCKLDQRTLYELHYQLITFGK---VFCTKSRPNACPMRGECRHFASAVASARLAL 1401
Db 1687 WSTLC---MLVELHTQ---GQDRFVY-----TCNECKKHVETRWHTCTVEDYDLCI 1731
QY 1402 PAPEERSLTSATIPVPPEFPFPAIPMIELPL-L-EKSLASGAPSNRENCEPIIEEPASP 1460
Db 1732 NCYNTKSHTH-----KMKWGLGLDDGSSQGEQPSK-----SP 1765
QY 1461 GQECFEIT-----ESDIEDAYYNEDPDEPTTKLNEIQFGMTLREHMERNMELQEG----- 1511
Db 1766 -QESRRLSIQRQISQILVHACQCRNANCSLPSC-----QMKRVVQHTKCKRK 1812
QY 1512 -----DMSKALVAL-----HPTTTSIIPP---KLK-NISRLRTEHOVYELPDSHRLDQ 1556

Db 1813 TNGGCVCKQJIALCCYHAKHCQENKCVFPCLNIKHNVRQOQIQHCLQOQAQLMRRMAT 1872

Qy 1557 MDKR---EPDDPSYLLAIWTPGE---TANSAQPEQ 1587

Db 1873 MNRNVPOSLPSISAPPGTPTQOPSTPQPPAQ 1909

RESULT 9

US-08-194-468-2

Sequence 2, Application US/08194468

Patent No. 5750336

GENERAL INFORMATION:

APPLICANT: Montminy, Marc R.

TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF

TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN

TITLE OF INVENTION: RESPONSIVE GENES

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark

STREET: 444 South Flower Street, Suite 2000

CITY: Los Angeles

STATE: California

COUNTRY: USA

ZIP: 90071

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/194,468

FILING DATE: 10-FEB-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Reiter, Stephen E.

REGISTRATION NUMBER: 31,192

REFERENCE/DOCKET NUMBER: P41 9672

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619)-546-4737

TELEFAX: (619)-546-9392

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2441 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-194-468-2

Query Match 2.0%, Score 183; DB 1; Length 2441;

Best Local Similarity 17.5%; Pred. No. 1.2e-05;

Matches 342; Conservative 222; Mismatches 685; Indels 708; Gaps 84;

Qy 5 MDSSAVNATEQNDGSRQVDFEFLNKT-----POKPSKRKRKPKMPKVYVGEKPKRK 59

Db 287 MGATGVNPQLASKQSMVNSLPFAFTDIKNTSVTPVNMQLQTSVGIPTQAIATGPTAD 346

Qy 60 P--RKPAELPKVYVGEKPKRKPKRAATQEKVKSETGSAKKNLKSAT-----KKPANVG 113

Db 347 PEKRLIQOQLVLLHLHAHKQRRQANGEVRACSLPHCRTMKNVNLNHHMTHCQAPKACOVA 406

Qy 114 DMSNKSPEVT-LKSCRK---ALNFDLENPGDARQGD-----SESEIVONSSGANSFSEIRD 165

Db 407 HCASSRQIISHKWKTRHDCVCLPLKNASDKRNQOQILGSPASGIQNTIGS-----V 459

Qy 166 AIGTNGSFLDSVSDIKTN-----GLGAMNQPLEV----- 196

Db 460 GAGQCNATSLSNPNPIDPSSMORAYAAALGPLYMNQPTQLQPVPGQQAQPPAHQQMRT 519

Qy 197 --SMGNQPKLSTCAKLARQDQPLLTRNQOCQFPVATONTQFPMENQQAQWLMKNOLIG 254

Db 520 LNALGNPMSPVAGG--ITTDQQPNLLISESALPTSLGATN---PLMND-----G 564

Qy 255 FPGNQOQRMRTIRNOQPCPLAMNQOPMY-----LIGTPRPAL----- 291

Db 565 SNSCNIGSLSTIPTAAPPSSTGVKRGWHEHVTQDLRSLVHKLVAIPTTDPDAALKDRR 624

Qy 292 -----VSGNQQLG 299

Db 625 MENLVAYAKKVEGDMYESANSRDEYHLLAEKIYKIQEKEKRRLHKGILGNQPAL 684

Qy 300 GPOGNKRPIFLNHOTCLPAGNOLYGSPTDMHOLVMSTGGQHQHLLIKNOQPGSLIRGOOP 359

Db 685 PASGAQPPVPPAQSVPVPPNGPL---PLPVNRMQVSG-----MNSFNPMISLGNVQLP 734

Qy 360 CVPLIDQOQPATPKGFTHLNMV--ATSMSS--EG--LRPHSQSQVPTTYLHVESVSRILNGT 415

Db 735 -----QAPMGPRASPMNHSVMNSMASVFGMAISPRMPQPPN-----MMGT 777

Qy 416 TGTQCRSRAPAYDSLQODIHOGNKYILSHEISNGCKKALPONSSLLPTPIMAKLEARG 475

Db 778 HANNIMAQAPTQNF-----LPQN----- 796

Qy 476 SKROYHRAMGOTEKHDLNLAOQIAQSODVERHNSSTCVE-----YLDAAKKTIKQKVVOE 530

Db 797 ---QFPSSGAMSVNSVGMQPAQAAGVSGQEPGAALPNPLNMLAFOASOLPCPPVTQS 853

Qy 531 NLHGMPPEVIEIDDP-----TDGARKGKNTASISKGASK 565

Db 854 PLHPTPPASTAACMPSLOHPTAPCMTPPOPAATQPTPVSSGOTPTPTPGSVPSAAQT 913

Qy 566 GNSSPVKKTAKEKCIYKPKTAKGRAGRKSKVPPPAHASEIQIWOPT-----PPKPTPLS 620

Db 914 QSTPTVQAAQAQVTPQPTPVQ-----PPSVATPOSSQOQPTPVHTQPPGTPLS 963

Qy 621 RSK-----PKGGRKSIQDSKARGPSGELLQODSTAETIYRMONLYLGDKREQEON 673

Db 964 QAAASIDNRVPTSTVTSATSSOQPGDPVPM-----EMKTEVQT-----DDAPE----- 1010

Qy 674 AMVLYKGDGALVPYESKRRKPRKVDIDDETTIRIWNLMGKDEKDEKDEKDEKDEKDEK 733

Db 1011 -----PTES--KGEPRSEMEED-----LQSSQVKEETDTTEQKSEPMEV 1049

Qy 734 EERRVFRGRADSFARMHLVOGDRRFPWKGVVDSVIGVFLTQNVSDHLSASSAFMSLAA 793

Db 1050 EEKK-----PEVKEAKEEEEEENSSNDTASQSTSP 1078

Qy 794 RFPPLSSSREDER-----NVRSVVVEDPEGCIL-----NLNIPSWQEKVOHPSDMEV 842

Db 1079 SQPRKKIKFPEELQALMPTLEALYRQDPESLPQRPVDPOLLGIPDFDVIKKNPMDLST 1138

Qy 843 --SGVDSGSKEQ-----LRDCSNSGIERNFLEKSIQNLEEEVLSSQDS 884

Db 1139 IKRKLDTGQYQEPWQYVDDVRLMENNALYNRKTSRV--YKFCSKLAEVFEQEI----- 1190

Qy 885 FDPAIFGCGRVGSCSKSDAEPTTTCETKTVSGTSQSVQTSQPNLSDBIC-----L 938

Db 1191 -DP-VMQS---LGYYCCGRKYEFSPTLCCYCKQJCTIPROAAYYSYQNRVHFCGKCFEI 1245

Qy 939 QGNERPHLYESGSDVQKQETTNVAQKPKDLEKTN----- 973

Db 1246 QG-ENVTL-----GDDPSQPTTIS--KDOPEKKKNDTLDPEFPVDCKECGKMKHQCIVLH 1298

Qy 974 ---WKDS--VC-----FGQPRNDTN-----QTTPSSSY-----EQCATRQPHVLIED 1012

Db 1299 YDIWPSGFVCDNCLKKTGRPRKENKESAKRLQTTLGNHLEDRVNFRLRRQNHEAGEV 1358

Qy 1013 F-----CMQGEGLGYSWMSISIPRVDRVKNKNVPR-----FFRQGGSVPR 1052

Db 1359 FVRVASSDKTVEYKPGMKSRFVDSGEMS-----ESFYRTKALFAFEIDGVDV 1408

Qy 1053 EFTQIILPST-----PHELPGM-----LSGSSSAVOE----- 1080

Db 1409 CFQGMHVQDALTAPHQIQCVIYISLDSTHFFRPRCLRTAVYHEILIGLYEVKLVYV 1468

Qy 1081 -----HQDD---TOHNQODEMNKASHLO-----KTFDLNLSSEECLETRQSS 1119

Db 1469 TAIWACPPSEDDYIFCHPDPQIPKPKRQEWYKKMLDKRAFAERIINDYKDFKQAN 1528
QY 1120 TKQNTIDGCLPRDRTAEDVVDPLSNSSNQNLIVESNSSNKQTAVEYKETTATILREMK 1179
Db 1529 -----EDRLTSAKELPYFGDFWPNVLESIRELEQEEERKKEESTAASETP 1577
QY 1180 GTLADGKPTSDSLRKDVENEGROENKNNMSIDYEAIRASISEISAIRER 1236
Db 1578 GSQGSK-----NAKKNNKTKNKK-----SSISRANKKPSM 1611
QY 1237 GMMNMLAVRIKDELER-----IVKDHGGIDLEWLRESPPDKADYLLSIRGLKSV 1289
Db 1612 PNVSDLSOKLATMEKKKEVFVHLHAG-----PVISTQPIVDPDPLLS----- 1658
QY 1290 CVRLTLHLNLAAPPVDTNVR-----IAYRMGWVPLQPLPESLQLHLELYPVLESTQKFL 1344
Db 1659 -----CDLMGRDAFLTLARDKH-----EFSILRSK 1686
QY 1345 WPLCKLQORTLYELHYLIITFGK-----VCTKSRNPNACPMRGCEGRHFASAYARLAL 1401
Db 1687 WSTLC-----MLVELHQ-----GODRFVY-----TCNECKHHVETRWHTCTVCEYDL 1731
QY 1402 PAPERSLTSATIPVPSPFPVAPIMIELPLPL-EKSLASGAPSNRENCEPIIEEPASP 1460
Db 1732 NCYNTKSTH-----KMKWGLGLDDEGSSQGEPOSK-----SP 1765
QY 1461 GOCEITEIT-----ESDIEDAYYNEDPEITIKLNEOFGMTLREHMERNMELQEG----- 1511
Db 1766 QESRRLSTQRCIQSLHACQCNANCSLPSC-----QMKRVVQHTKCKRK 1812
QY 1512 -----DMKALVAL-----HPTTSITPP-----KUK-NISRLRTHQVVELPDSHRLD 1556
Db 1813 TNGGCPVCKQLLALCYHAKHCQENKCPVPFCLNKHNVROQOIQHCLQQAQLMRRMAT 1872
QY 1557 MDKR-----BPDDPSYLLAIWTPGE-----TANSQAQPP 1587
Db 1873 MTRNVPOQSLSPTSAPPGTPTQPTPQPPAQ 1909

RESULT 10

US-08-598-591-2
; Sequence 2, Application US/08598591
; Patent No. 5654155
; GENERAL INFORMATION:
; APPLICANT: Allen, Antonette C.
; APPLICANT: Alvares, Christopher P.
; APPLICANT: Critz, Brenda S.
; APPLICANT: Murphy Patricia D.
; APPLICANT: Olson, Sheri J.
; APPLICANT: Scheiter, Denise B.
; APPLICANT: Zeng, Bin
; TITLE OF INVENTION: A Consensus Sequence of the Human BRCA1 Gene
; Patent No. 5654155
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS
; STREET: 699 Prince St.
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/598.591
; FILING DATE: herewith
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Swecker, Robert S.

; REGISTRATION NUMBER: 19,885
; REFERENCE/DOCKET NUMBER: 020160-282
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6620
; TELEFAX: 703-836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: BRCA1
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 17
; MAP POSITION: 17q21
; US-08-598-591-2

Query Match

Best Local Similarity 2.0%; Score 182; DB 1; Length 1863;

Matches 293; Conservative 195; Mismatches 494; Indels 608; Gaps 72;

QY 20 DGSRODVLEF---DLNKTPQPKSKRRKFMKPVVVEGPKR----- 58
Db 219 DSAKKAACEFSDVTNTEHQPSNNDLNTTEKRAAERHPEKYQGVSSVSNLHVEPCGTNT 278
QY 59 -----KPKPAELPKVVVEGPK-----KPKPAELPKVVVEGPK 98
Db 279 HASSLOHENSLLLTQDRMNVEKAEFCNKSQKOPGLARSOHNRWAGSKETCNDRRTPTSEK 338
QY 99 K-----NKESATKKPANYGDMNSKSPVTLKSCRKALN-----FDLEN 137
Db 339 KVDLADPLCERKENKQKLPCESENPROTEDEV---PWITLASSIOKVNEWFSRDEL 394
QY 138 PGDARQGSSEI-----VONSSGANS-----FS 161
Db 395 SDDSHDGESENAKADVLDVLNEVDEYSGSEKIDLLASDPHEALICKSERVHSKSVES 454
QY 162 EIRDAIGTGN-----GSLDSVSQIDK---TNGLGAMNQPLEVSMG 199
Db 455 NIEDKIFKTYRKASLPLNLSHVTEENLIIGAVFTPEIIQERPLTNKLRKRRP---TSG 511
QY 200 NOPDKLSTGAKLARDQPDLLTR---NOQCQFVATQNTQFPMENOQAWLQMKNLIGPFP 257
Db 512 LHPEDFIKKADLAVOKTPEMINQGTQEQNGQVMNITNSGHENKTGDSIQNE----- 565
QY 258 GNQQRWTRNQOQCLAMGNQOQPMYLIPTPRPALVYGNQOQGLGPGQGNKRPIFLN-HOTCL 316
Db 566 -----KNPNIESLEKESAFKTAEPISISSIS-NMELE-----LNTHNSKA 605
QY 317 PAGNOLYGSPTDMH---OLVMS-----TGGOHGLLIKNOQPGSLIRG 356
Db 606 PKNRLRRKSSSTRHIALELVVSRNLSPNPNCTELQIDSCSSSEETKKKYNOMP---VRH 662
QY 357 QOQPCVPLIDQOPAT-----PKGFT-----HLNQMV 381
Db 663 SRNLQIMECKEPATGAKKSNKNEQTSKRHSDTTPPELKLITNAPGSFTKCSNTSELKEPV 722
QY 382 ATSM-----SSPGLRPHSQSOVP---TTLHVES 407
Db 723 NPSPREEKEELETVKVSNNAEDPKDMLSGERVLTQERSVSSISILVPGTDYGTQES 782
QY 408 VSRILNGTGTQOR-----SRAPAYDSIQODIH-----OGNKYILSHSINGNG 451
Db 783 ISLLEVSTLTKAKTEPNKCVSOCAAFENPKGLIHGCKDNDRNDTEGFKYPLGHEV----- 837
QY 452 CKKALPQNSSLPTIMAKLEEA-----RGSKRO-----YHRA 483
Db 838 -----NHSRETSI--EMESELDAQVLYQNTFKYKRSQSFALFSPNPGNAEECATFS 888

```

484 MGQTEKDLNLAQIAQSDQVHRHNSSTCEVYLDAAKTKTIQKVQENL-HGMPPEVIEI 542
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
889 SGLKKQSPKVTCEQKEENQKNES-----NIKPVQVNTITAGFP--VVGQ 934
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
543 EDDPTDQAR---KGKNTASTS---KGASKNSPVPKTAKEKCIYP-----KTPA 587
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
935 KDKPVDNAKCSIRKGSFRCLSSQFRNETGLITPNKHGLLQNPYRIPPLFPKSFVKTKC 994
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
588 KKGKGRKKSVPPPAHASAETQLWQTPPKT--PLSRSKP-----KKGKRSIODSGKARG 640
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
995 KKNL---EENFEHNSPEHEMGENTIPSTVTSIRNIRENVFKGASSNINEVGSSTN 1052
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
641 PSGBELLQDSTAEIIRMQNLY--LGBKEREQONAMVLYKGDALVPYESKK----- 691
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1053 EVG-----SINEIGSSDENIAELG--RNRGPKLNAMRL--GVLOPEYVKSLPSCSNC 1103
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
692 KRPKPKVIDDETRINWL-----LGMKGDEKEGD 721
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1104 KHPEIKKQEEYEVQVNTDFSPYLISDNLEQPMGSSHASQVCSETPDDLDDGEIKEDT 1163
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
722 E--EKDKKKKEKWEERRVP-----RG---RADSFIAHMLVQDREFFSPWKGSVVDSVI 771
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1164 SFAENDIK-----ESAVFSKSVQRELSRSPFTHHLAQYRGA----- 1206
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
772 GVFLTONVSDHLSSAPMSLAARFPFKLSSSRREDERNVRSVVVWEDPE--GCILML-----N 825
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1207 -----KKLESS-----EENLSS---EDELPCFQHLFGKVN 1235
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
826 EIPSWQKVOHPDSMEVSGVDSGSKE-----QLRDCNSGIERFNFLKSTQNLEEV 878
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1236 NIPS--QSTRH-STVAPECLSKNTEENLLSLKNSLNCNOVI-----LAKASQ---EHH 1284
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
879 LSSQDSFDPALFOSGRCRGVSCSCSDAEPTTRCETKTVSGTSOSVQTGSPN---LSD 934
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1285 LSETKCSALFSS-----QSELEDLTANTQDPFLIGSKOMRHOSESGCVGLSD 1337
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
935 EICLQGNRPHLYEGSDVQKQETTNVAQKPDLEKTMNMKDSVCFGQPRNDTNWQTPS 994
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1338 KELVSDDEE---RGTG---LEENN--QEEQMSDNLGEAASGC-----ESET 1376
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
995 SSVYQCATRQPHVLDIDFCMQEGGLGYSNWSISPRVDRVKNKNVPRFRFGSGVPREF 1054
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1377 SVSEDCS----- 1383
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1055 TGOIIPSTHELPGMGLGSSSAVQEHQDIT--QHNQODENMKASHLOKTFDLNLSSEEC 1113
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1384 -----GLSSQSDILTTQORTMQHN-----LIKQOEMAELEAVLEQH 1421
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1114 LTROSSTKQNTDGCUPRDRFTAEDVDPLSNSSQLNVLVSNSSKNKEQTAVYEKTNAT 1173
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1422 GSQPSNSYPSI-----ISDSSALEDLRNPQESTSEKAVLTSQKSSEYP 1464
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1174 ILREMGKTLADGKKPTQOWDSLKRDVEGNGRGRERNKNMDSIDYAIRASISEISEAI 1233
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1465 ISONPEGLSAD--RFEVSADSSSTKNKE--PGVERSFSKPCPSLD-----DRWMHSCSGSL 1517
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1234 KERGM--NMMLAVRIKDFLERIVKDHGGIDL 1262
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1518 QNRNYSQEEELIKVVDVEEQLEESGPHDL 1547
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 11

```

US-08-798-691-2
; Sequence 2, Application US/08798691
; Patent No. 5750400
; GENERAL INFORMATION:
; APPLICANT: Murphy, Patricia D.
; APPLICANT: Allen, Antonette C.
; APPLICANT: Alvares, Christopher P.
; APPLICANT: Critz, Brenda S.
; APPLICANT: Olson, Sheri J.
; APPLICANT: Schelter, Denise B.
; APPLICANT: Zeng, Bin

```

```

; TITLE OF INVENTION: Coding Sequences of the Human
; TITLE OF INVENTION: BRCA1 Gene
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ONCORMED
; STREET: 200 Perry Parkway
; CITY: Gaithersburg
; STATE: MD
; COUNTRY: USA
; ZIP: 20877
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/798,691
; FILING DATE: 12-Feb-97
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Thomas Gallegos
; REGISTRATION NUMBER: 32,692
; REFERENCE/DOCKET NUMBER: PA-0054CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-527-2051
; TELEFAX: 301-208-6997
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1863 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; STRAIN: BRCA1
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: 17
; MAP POSITION: 17q21
;
US-08-798-691-2

```

Query Match 2.0%; Score 182; DB 1; Length 1863;

Best Local Similarity 18.4%; Pred. No. 8.9e-06;

Matches 293; Conservative 195; Mismatches 494; Indels 608; Gaps 72;

```

QY 20 DGSRDVLEP---DLNKTPOOKPSKRRKPKMPKVVEGPKR----- 58
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 219 DSAKKAACEFSETDVTNTEHHQPSNNDLNTTEKRAAERHPKRYOGSSVSNLHVPCGTNT 278
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 59 -----KPRKPAELPKVVVEGKP-----KRPKKAATQEKVKKSGTSAKK 98
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 279 HASSLOHENSLLLTKDRMNVKAEFCNKSQKPLARSOHNRWAGSKETCNDRTPTSTEK 338
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 99 K-----NLKESATKKPANVGDMSNKGPEVTLKSCRKALN-----FDLEN 137
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 339 KVDLNADPLCKERKWNKQLPCSENPRDTEV---PWITLSSSIQKVNFWFSRDELLG 394
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 138 PGDARQGDSESEI-----VQNSSGANS-----FS 161
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 395 SDDSHDGESESNAKVADVLDVLNEVDSEYSGSEKIDLLASDPHEALICKSERVHSKSVES 454
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 162 EIRDAIGGTTN-----GSFLDSVSQIDK---TNGLGAMNOPLEVSMG 199
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 455 NIEDKIFGKYRKKASLPNLSHVNTENLLIGAFVTEQIIQERPLTNKLRKRRP---TSG 511
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 200 NOPDKLSTGAKLARDQOQDPLLTR--NQOQFPVATQNTOPPMENQQAOWLMQNLIGFPF 257
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 512 LHPEDEFIKKADLAVQKTPTEMINQNTQTEQNGVMNITNSGHENKTKGDSIQNE----- 565
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 258 GNOQPRMTIRNQOPLAMGNOQPMYILGTPRALVSGNOOLGGPOGNKRPIFLN--HOTCL 316
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 566 -----KNPNPIESLEKESAFKTKAEPISSIS--NMELE-----LNIHNSKA 605
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

[illegible]

Db 279 HASSLOHENSLLLT KDRMNVKAEFCNKSKQPLARSOHNRWAGSKETCDNRRTSPTEK 338
QY 99 K-----NLKESATKKPANVCDMSNKSPEVTLKSCRALN-----FDLEN 137
Db 339 KVDLNADPLCEKWNKQLPCSENPRDTEV-----PWITLNSIOKVNEWFSRDELIG 394
QY 138 PGDARQDSESEI-----VONSSGANS-----FS 161
Db 395 SDOSHGESNESNAKADVLDVNEVEDEYSGSSEKIDLLASDPHEALICKSERVHSKSVES 454
QY 162 EIRDAIGGTH-----GSFLDSVSQIDK-----TNGLGAMNOPLEVSVMG 199
Db 455 NIEDKIFGTYRKASLIPNLSHVTENLIIGAFVTEQIIQOERPLTNKLRKRPP-----TSG 511
QY 200 NOPDKLSTGAKLARDQOPLLTR--NOQOFPVATONTOPPMENOQAWLQKOLGPF 257
Db 512 LHPEDFIKKADLAVQKTPENINGOTNQTONGOVAMNITNSGHENKTKGDSIQNE----- 565
QY 258 GNOQPRMTIRNOQPCLAMGNQOQPMYLGTPRALVSGNOQLGPGQGNKRPIFLN--HOTCL 316
Db 566 -----KNPNPIESLEKESAFKTAEPISISSIS--NMELE-----LNIHNSKA 605
QY 317 PAGNQLYGSTDMH-----QIVMS-----TGGQOHLGKILKNOQPGSLIRG 356
Db 606 PKNLRKRSSTRIHALELVWNRNLSPNCTELQIDSCSSSEIKKKKYNOMP-----VRH 662
QY 357 QOPCVPLIDQOPAT-----PKGFT-----HLNQMV 381
Db 663 SRNLQMEGKEPATGAKSKNPNQTSKRHSDTFFPELKLITNAPGFTKCSNTSELKEFV 722
QY 382 ATSM-----SSPLRPHSQSQVP--TYTLHVES 407
Db 723 NPSLPREEKEKLETVKVSNNAEDPKDLMSLGERVLQTERSVESSISLVPDGYGTQES 782
QY 408 VSRILNGTTCOR-----SRAPATVDSLODTH-----QGNKYLSHEISNGNG 451
Db 783 ISLEVSTLGAKTEPNKCVSQAAPENKGLIHGCSKDNRNDETEGFKYPLGHEV----- 837
QY 452 CKKALPONSLPTPIAKLEEA-----RGSKRQ-----YHRA 483
Db 838 -----NHSRETSI--EMESELDQALONTFKVSKRQSFALFSPGNAEECATFSAH 898
QY 484 MGQTEKHDNLQAQSQDQVERHNSSTVEYLDAAKTKIKVQENL--HGMPPEVIEI 542
Db 889 SGSLKQSPKVTCEQKEENOCKNES-----NIKPVQTVNITAGFP--VVGQ 934
QY 543 EDDPTDGR--KGKNTASIS--KGASKNSSPVKTAKEKCIVP-----KTPA 587
Db 935 KKPVDNAKSIKGGSRFCUSSQFRNETGLITPNKHLGLONRYRIPPLPKSFVKTKC 994
QY 588 KGRAGRKKSVPPPAHASEIQLMQTPPKT--PLSRSKP-----KKGKRSIQDSGKARG 640
Db 995 KKNLL--EENFEHSMSPEREMGNENIPSVTISRNNIRENVFKGASSNINEVGSSTN 1052
QY 641 PSELGQDSIAEIIYRMQNL--LGDKEREQONAMVLYKGDGALVPEYSK----- 691
Db 1053 EVG-----SSINEIGSDENIQALG--RNRGPKLNLMLR--GVLQPEYKQSLPGSNC 1103
QY 692 RKPVPKVIDIDETTRIWL-----LMGKGDEKEDG 721
Db 1104 KHPKIKQEEYEVVQVNTDPSLYLSDNLEQPMQSSHASQVSETPDDLLDDGETKEDT 1163
QY 722 E--EKDKKKEKWEERRVF-----RG--RADSFIAHMLVQDGRFSPWKGSVVDSVI 771
Db 1164 SFAENDIK-----ESSAVFESKSVQRGELSRSPSPHTHLAGYRGA----- 1206
QY 772 GVFLTONVHLSSAFMSLAARPPKLSRREDENRVSVVDEPDE--GCILNL-----N 825
Db 1207 -----KLESS-----EENLSS--EDELPCFQHLHFGKVN 1235
QY 826 EIPSWQEKVQHPDMEVSGVDSKSE-----QLRDCNSNGIERNFLEKSTONLEEEV 878
Db 1236 NIPS--OSTRH--STVATECLUSKNTENLJLSKLNLSNDCSNQVI-----LAKASQ-----EHH 1284

QY 879 LSSQDSFDPALFOSCGRVGSCSKSDAEPTTRCTETKTVSGTSQSVQTSQPN-----LSD 934
Db 1285 LSEETKCSASFSS-----QCSELEDITANTNTODPFLIGSSKQMRHOSQGVCLSD 1337
QY 935 EICLOGNERPHLYEGSDVQKQETTINVAQKKPDKLEKTMNMKDSVCFGQPRNDTNWQTTPS 994
Db 1338 KELVSDDEE--RGTG--LEENN--QEEQMSDNLGEAASGC-----ESET 1376
QY 995 SSVTEQCATROPHVLDIEDFCMQEGEGLYSWMSSISPRVDRVKNKNVPRFRFQSGSVVPREF 1054
Db 1377 SVSEDCS----- 1383
QY 1055 TGOIIPSTHELPCMGLSGSSSAVQEHODDT--OHNOODEMKNKASHLOKTFDLNLSSEEC 1113
Db 1384 -----GLSSQSDILITQORTMQRN-----LIKLOQEMAELEAVLEQH 1421
QY 1114 LTRQSTKQONITGCLPRDRTAEDVVDPLSNNSLQNLVSNSSNKKEOTAVYKETNAT 1173
Db 1422 GSQPSNSYPSI-----ISDSSALEDLRNPQESTSEKAVLTSQKSSEYP 1464
QY 1174 ILEMKGTLADGKKPTSQWDSLRKDYEGNEGRQERNKNMDSIDYEAIRRASISEISEAI 1233
Db 1465 ISONPEGLSAD--RFEVSADSSTSKNKE--PGVERSSPKCPSLD-----DRWYMHSCSGSL 1517
QY 1234 KERGM--NNMLAVRIKDFLERIVKHGIGIDL 1262
Db 1518 QNRYPSQBELIKVVDVEEQLEESGPHDL 1547
RESULT 13
US-08-825-487A-2
: Sequence 2, Application US/08825487A
: Patent No. 6048689
: GENERAL INFORMATION:
: APPLICANT: Murphy, Patricia D.
: APPLICANT: White, Marga B.
: TITLE OF INVENTION: METHODS FOR IDENTIFYING VARIATIONS IN POLYNUCLEOTIDE SEQUE
: NUMBER OF SEQUENCES: 110
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Howrey & Simon
: STREET: 1299 Pennsylvania Avenue., N.W.
: CITY: Washington,
: STATE: DC
: COUNTRY: USA
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/825,487A
: FILING DATE: 28-MAR-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US98/060002
: FILING DATE: 26-Mar-1998
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Albert P. Halluin
: REGISTRATION NUMBER: 25,227
: REFERENCE/DOCKET NUMBER: 05371.0012.999
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-463-8100
: TELEFAX: 650-463-8400
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1863 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: protein

ORIGINAL SOURCE:
ORGANISM: Homo sapiens
STRAIN: BRCAL
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 17
MAP POSITION: 17q21
US-08-825-487A-2

Query Match 2.0%; Score 182; DB 3; Length 1863;

Best Local Similarity 18.4%; Pred. No. 8.9e-06;

Matches 293; Conservative 195; Mismatches 494; Indels 608; Gaps 72;

QY 20 DCSRDVLEF---DLNKPQPKRKRKFKMPKVVEKPKR-----58
DB 219 DSAKKAACEFSDTNTTEHHQPSNDLMTTEKRAAERHPKRYGSSVNLHVEPCGTNT 278
QY 59 -----KPKPAELPVVVEGKP-----KPKPKAATOEVKSKETGSARK 98
DB 279 HASSLOHENSLLLTQDRMNVKAEFCNKSQOPGLARQHNWAGSKETCDRRTPSTEK 338
QY 99 K-----NLKESATKPKANVGMNKSPEVTLKSCRKALN-----FDLEN 137
DB 339 KVDLNADPLCERKENKQKLPCESENPRTEDEV---PWITLASSIQKYNWESRSDLELLG 394
QY 138 PGDARQGSSEI-----VONSSGANS-----FS 161
DB 395 SDDSHDGESESNAKVADVLDVLENDVEYSGSEKIDLLASDPHEALICKSERVHSKSVEBS 454
QY 162 ETRDAIGTNN-----GSFLDSVSQIDK---TNGLAMNQPLEVSG 199
DB 455 NTEIDIFGKTYRKXKASLPNLHVNTENLIAGAVTEPQIIIOERPLTNKLRKRKP---TSG 511
QY 200 NOPDKLSTGAKLARDQPDLLTR--NQOCQFPVATQNTQFPMENOQAMLOMKNQLIGPFP 257
DB 512 LHPDEFIKKADLAVOKTPEMINQNTQTEONGOVNITNSCHENTKGDSIONE-----565
QY 258 GNOQPRWIRNOQPLANGNOOPMYLIGTPRALVYSGNOQGGPOGNKRPIFLN--HOTCL 316
DB 566 -----KNPNPTELESKEFAFKTAEPISSSIS--NMELE-----LNHNSKA 605
QY 317 PAGNOLYGSPTDMH---OLVMS-----TGQOQHGLLKNQOQSLIRG 356
DB 606 PKNLRKRSSTRHIALELVVSRNLSPPNCTELQIDSCSSSEIEKKKYNQMP---VNH 662
QY 357 QOCPVPLIDQOPAT-----PKGFT-----HLNOMV 381
DB 663 SRNLQIMEGKEPATGAKKSNKPNEQTSKRHSDTTPPELKLTPAGSFTKCSNTSELKEV 722
QY 382 ATSM-----SSPGLRPHSQSQVP--TTYLHVES 407
DB 723 NPSLPREEKEELETVKVSNNAEDPKDMLSGERVLTQTERSVESSISILVPGTDYGTQBS 782
QY 408 VSRILNGTGTQCR-----SRAPAYDSLOQDIH-----QGNKYILSHEISNGNG 451
DB 783 ISLEEVSTLGKAKTBNKCVSOCAAFENPKGLIHGCSKDNNDTEGFKYPLGHEV-----837
QY 452 CKKALPQNSLPTPIAKLEEA-----RGSKRO-----YHRA 483
DB 838 -----NHSRETSI--EMEESLDAQYLQNTFKVSKRQSPALFSPNGNAEECATFSAH 888
QY 484 MQOTEKHLNLAAQTAQSDQVDRHNSSTCVYLDAAKTKIKQVVOENL--HCMPPEVIEI 542
DB 889 SGSLKQSPKVTFECEQKBENOGKNES-----NIKPVQTVNITAGFP--VVGQ 934
QY 543 EDDPTDGAR---KGNTASIS---KGAGNSSPVKKTAEKEKICVP-----KTPA 587
DB 935 KDKVDNKAISLGGSRCLSSQFQNGTGLTIPNKHGLLQNPYRIPPLPKSFVKTKC 994
QY -588 KKGAGRKKSVPVPAHASEIQWLQPTTPKPT---PLSRSKP-----KGKGRKSQDQSGKARG 640
DB 995 KKNLL--EENFEHNSMSPREMGNEINIPSTVSTISRNNIRENVFKGASSNINEVGSSTN 1052

QY 641 PSCELLCQDSIAEILIRMONLY--LGDKEREQEQNANWLYKGDGALVPYESKK-----691
DB 1053 EVG-----SSINEIGSDENIOAELG--RNRGPKLNAMLRL---GVLOPEVYKOSLPGSNC 1103
QY 692 RKPRKVIDDDETTIRIWL-----LMGKGDEKEGD 721
DB 1104 KHEPKIKOEYEVEVQTVNTDFSPYLLISDNLEQPMGSSHASQVCSETPDLDLDDGEIKEDT 1163
QY 722 E--EKDKKKKEKWEERRVF-----RG-----RADSFARMHLVQDGRFRFPKSGVVDVSI 771
DB 1164 SPAENDIK-----ESSAVFSKSVQGELSRSPSPFTHLAQGYRGA-----1206
QY 772 GVFELQNSDHLSSAFNSLAARFPKLLSSSRREDERNVSVVVEDEPE--GCILNL-----N 825
DB 1207 -----KKLESS---EENLSS-----EDEELPCFOHLLFGKYN 1235
QY 826 EIPSWOEKVOHPDMEVSGVDSGSKE-----QLRDCNSNGIERFNFLEKSIQNLEEV 878
DB 1236 NTPS--QSTRH--STVATECLSKNTEENLLSKNSLNDSCNOVI-----LAKASQ---EHH 1284
QY 879 LSSQSDFPAPFQSCGRVSGSCSKSDABFPTRTCTKTVSGTSOSVOTGSPN---LSD 934
DB 1285 LSEETKCSASLFSS-----QCSELEDLTANTQDPFLIGSSKOMRHOSESQGVGLSD 1337
QY 935 ETCLOGNERPHLYEGSDVQKQETTINVAQKPDLEKTMNWKDSVCEGQPRDNTNWTTPS 994
DB 1338 KELVSDDEB---RGTG---LEENN--OEQMSDNLGEAASGC-----ESET 1376
QY 995 SSYEOCATRQPHVLDIEDFGMOGEGLYGWSMSISPRVDRVKNKNVPRFRFROGGSVPREF 1054
DB 1377 SVSEDCS-----1383
QY 1055 TQGIIPSTPHELPGMLGSSSAVOEHQDDT--QHNQODEMNKASHLQKTFDLNLSSEC 1113
DB 1384 -----GLSQSDILITQQRDTMOHN-----LIKLOEAMAELEAVLEQH 1421
QY 1114 LTRQSTKQNIITDGLPRDRTAEDVVDPLSNSSNLONILVESNSSNKEQTAVEYKRETNAT 1173
DB 1422 GSQPSNYSPI-----ISDSALEDLRNPQSTSEKAVLTQSKSEYP 1464
QY 1174 ILREMGKTLADCKKPTQOWDSLKRDVEGNEGQERKNMNDSDIYAIRRASITSEIAI 1233
DB 1465 ISQNEGLSAD--KFEVADSSTSKNKE--PGVERSPSKCPSLD-----DRWYHSCSGSL 1517
QY 1234 KERGM--NNLAVRIKDFLERIVKDHGGIDL 1262
DB 1518 QNRNYPQOEELIKVYDVEEQLEESGPHDL 1547

RESULT 14

US-08-825-487A-6

; Sequence 6, Application US/08825487A

; Patent No. 6048689

; GENERAL INFORMATION:

; APPLICANT: Murphy, Patricia D.

; APPLICANT: White, Marga B.

; TITLE OF INVENTION: METHODS

; NUMBER OF SEQUENCES: 110

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Howrey & Simon

; STREET: 1299 Pennsylvania Avenue., N.W.

; CITY: Washington,

; STATE: DC

; COUNTRY: USA

; ZIP: 20004

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/825,487A

; FILING DATE: 28-MAR-1997

FOR IDENTIFYING VARIATIONS IN POLYNUCLEOTIDE SEQUE

CLASSIFICATION: 435
PRIOR APPLICATION DATA: PCT/US98/060002
FILING DATE: 26-Mar-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Albert P. Halluin
REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: 05371.0012.999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-463-8100
TELEFAX: 650-463-8400
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
STRAIN: BRCAL
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 17
MAP POSITION: 17q21
US-08-825-487A-6

Query Match 2.0%; Score 182; DB 3; Length 1863;
Best Local Similarity 18.4%; Pred. No. 8.9e-06;
Matches 293; Conservative 195; Mismatches 494; Indels 608; Gaps 72;
QY 20 DGSQDVLEF---DLNTPQOKPSKRRKMPKVVVGPKPKR-----58
DB 219 DSARKACESETDVTTEHQPPNSNDLNTKRAAERHPEKYQGVSSVNLHVEPCGTNT 278
QY 59 -----KPRKPAELPKVVVEGKP-----KRKPRKAATOPKVKSKETGSAKK 98
DB 279 HASSLOHENSLLLTDRMNVKAEFCNKKQPGKLARSQHNWAGSRETCDNRTPSTEK 338
QY 99 K-----NLESATKKPANVGMNSKSPVTLKSCRKALN-----FDLEN 137
DB 339 KVDLNADPLCKERWKKQKLPCESENPRDTEV---PWITLNSIQKVNWFNSRDELLG 394
QY 138 PGDARQDSESEI-----VQNSGANS-----FS 161
DB 395 SDOSSHGSESENAKADVLDVLENEVDEYSGSEKIDLLASDPHEALICKSERVHKSVES 454
QY 162 EIRDAIGTN-----GSFLDSVSQIDK-----TNGLGAMNOPLEVSMG 199
DB 455 NIEDKIFGKYRKKASLPNLSHVNTENLIIGAFVTEPQIIQERPLTNKLRKRRP---TSG 511
QY 200 NOPDKLSTGAKLARDQPDLLTR--NQCCQFPVATQNTQFPMENOQAWLQKQLIGFPF 257
DB 512 LHPEDFTFKADLAVQKTPENINQGTNQTQEQVMNITNSCHENKTKGDSIQNE-----565
QY 258 GNOQPMRTINQOPLCANGQOQMYLLIGTRPALVSNQOGLGGPQGNKRPIELN-HOTCL 316
DB 566 -----KNPNPIESLEKSAFKAEPISSIS-NMELE-----LNHNSKA 605
QY 317 PAGNQLYGSPTDMH-----QLVMS-----TGGOHQGLLILKNQOPGSLIRG 356
DB 606 PKNRLRRKSSTRHIALELVSNRNLSPNPNCTELQIDSCSSSEIEKKKYNQMP---VRH 662
QY 357 QQCVPILDOQAT-----PKGFT-----HLNQMV 381
DB 663 SRNLQMEGKEPATGAKSKNPNQTSKRHSDTFPELKLITNAPGSFTKCSNTSELKEFV 722
QY 382 ATSM-----SSPLRPHSQSOVP-TTYLHVES 407
DB 723 NPSPREEKEKELETVKVSNNAEDPKDMLSGRVLQTERSSESSISLVPDGTGYTOES 782
QY 408 VSRILNCTTCTCOR-----SRAPAYDSLOQDIH-----QGNKYILSHEISNGNG 451

DB 783 ISLEVSTLCKAKTEPNKCVSOCAAFENPKGLIHGCKSDNRNDTEGFKYPLGHEV-----837
QY 452 CKKALPONSSSLPTPIAKLEEA-----RGSRRQ-----YHRA 483
DB 838 -----NHSRETSI---EMBESELDQYLONTFKVKRQSFALFSPNGNAEEECATSAH 888
QY 484 MGOTEKHDNLAAQIAQSDQVERHNSSTCVLEYLDAAKTKTIQKVVOENL-HGMPPEVIEI 542
DB 889 SGLKQSPKVPTECEKEENOQKNES-----NKKPVQTVNITAGFP--VVGQ 934
QY 543 EDDPTDGR---KGKNTASIS---KGASKNSSPVKTAKEKCIVP-----KTPA 587
DB 935 KDKPVDNAKCSIRKGSFRCLSSQFRNETGLITPNKHLQNLQNPYRIPPLPKSFVTKC 994
QY 588 KKGKGRGKSVPPPAHASEIOLWQTPPKT--PLSRSKP-----KGGKRSIODSGKARG 640
DB 995 KKNLL--EENFEHMSPEREMENIPSTVSTISRNIENRVFKGASSNINEVGSTN 1052
QY 641 PSGELLCQDSIAEIIYRMQNLV--LGDKEREQONAMVLYKGDGALVPYESKK-----691
DB 1053 EVG-----SSINEIGSSDENIQAEELG--RNRGPKLNAMRL--GVLQPEVYKQSLPGSNC 1103
QY 692 RKPRPKVIDDDETRINWL-----LMKGDEKEGD 721
DB 1104 KHPEIKKQEEYEVVQTVNTDFSPYLIISDNLEQPMGSSHASQVCSETPDDLDDGEIKEDT 1163
QY 722 E--EKDKKKKEKWEERRVF---RG---RADSFIAARMHLVQGDREFSPWKGSVVDVSI 771
DB 1164 SPFENDIK-----ESSAVFSKSVQGELSRSPFPFTHLAQYRGA-----1206
QY 772 GVFLTONVSDHLSSAPMSLAARFPFKLSSRREDERNVRSVVVEDPE-GCILNL-----N 825
DB 1207 -----KKLESS---EENLSS---EDELPCFQHLFLGKVN 1235
QY 826 EIRSWQEKVOHPDSMEVSGVDSSKE-----QLRDCSNSGIERFNFLEKSTONLEEV 878
DB 1236 NIPS--OSTRH-STVAPECLSKNTEENLLSKNSLNDSCNQVI-----LAKASO---EHH 1284
QY 879 LSSQDSFDPAIFOSCGRVGSCSDAEFPPTTTCETKTVSGTSQSVOQTGSPN---LSD 934
DB 1285 LSEETKCSALFSS-----QCSELEDLTANTQOPFLIGSKQMRHSESGOGLSD 1337
QY 935 EICLQGNRPHLYEGSGDVQKQETTNVAAKKPDLEKTMNMKDSVCFQGPQRNDTNWQTTPS 994
DB 1338 KELVSDDEE---RGTG---LEENN--QEEQSMDSNLGEAASGC-----ESET 1376
QY 995 SSTEQCATRQPHVLDIEDFCMQGEGLYSWMSSISPRVDRVKNKNVPRFRFQGGSVVPREF 1054
DB 1377 SVSEDCS-----1383
QY 1055 TGOIIPSTPHELPGMGLSGSSSAVQEHODDT-OHNQODENMKASHLOKTFDLDLNLSSEEC 1113
DB 1384 -----GLSSQSDILITTOORDTMQHN-----LIKQOEMAELEAVLEQH 1421
QY 1114 LTRQSTQKQITDGLCLPRDRTAEDVVDPLSNSSLNILVESNSSNKEQTAVYEKTNAT 1173
DB 1422 GSQPSNSYPSI-----ISDSSALEDLRNPQESTSEKAVLTSOKSSEYP 1464
QY 1174 ILREMGKTLADGKKPTQOWDSLRKDVEGNEGRNKNMNSIDYEAIRASISEISEAI 1233
DB 1465 ISONPEGLSAD-RFEVSADSTSKNKE--PGVERSSPKCPSLD-----DRWMHSCSGSL 1517
QY 1234 KERGM-NMMLAVRIKDFLERIVKDHGGIDL 1262
DB 1518 QNRNYPSEELIKVVDVEEQOLEESGPHDL 1547

RESULT 15
US-09-074-476-2
; Sequence 2, Application US/09074476
; Patent No. 6130322
; GENERAL INFORMATION:

APPLICANT: Murphy, Patricia D.
APPLICANT: Allen, Antonette C.
APPLICANT: Alvares, Christopher P.
APPLICANT: Critz, Brenda S.
APPLICANT: Olson, Sheri J.
APPLICANT: Thurber, Denise
APPLICANT: Zeng, Bin
TITLE OF INVENTION: Coding Sequences of the Human
TITLE OF INVENTION: BRCAL Gene
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howrey & Simon
STREET: 1299 Pennsylvania Avenue N. W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/074.476
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/074.453
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Albert P. Halluin
REGISTRATION NUMBER: 25,227
REFERENCE/DOCKET NUMBER: 5371.34.US01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-463-8109
TELEFAX: 650-463-8400
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1863 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
STRAIN: BRCAL (om11)
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 17
MAP POSITION: 17q21
US-09-074-476-2

Query Match 2.0%; Score 182; DB 4; Length 1863;

Best Local Similarity 18.4%; Pred. No. 8.9e-06;

Matches 293; Conservative 195; Mismatches 494; Indels 608; Gaps 72;

Qy 20 DGSRODVEF---DLNKTPOOKSPKRRKFKPVVVEGPKR-----58
Db 219 DSAKKAACEFSETDVTNTEHHQPSNNDLNTTEKRAAERHPKRYQSSVSNLHVPCGTNT 278
Qy 59 -----KPKPAELPKVVVEGKP-----KPKPKAAATQKVKSKETGSACK 98
Db 279 HASSLOHENSLLLTQDMNVKAEFCNKSQKQGLARSOHNRWAGSKETCDNRTPSTEK 338
Qy 99 K-----NLKSAKPKPANVGMNSKSPVTLKSCRKALN-----FDLEN 137
Db 339 KVDLADPLCERKENNKKLPCSENPRTEV-----PWTLNSSIQKYNWFSRDELLG 394
Qy 138 FCDARQDSESEI-----VQNSSGANS-----FS 161
Db 395 SDDSHDGESENAKVADVLDVLINEYDEYSGSEKIDLLASDPHEALICKSERVHSKSVES 454
Qy 162 EIRDAIGTGN-----GSPFLDSVSQIDK---TNGLGANNQPLEVSMG 199

Db 455 NIEDKIFGKYRKASLPNLNLSHTVNTLIGAFVTPBQIIQERPLTNKLRKRPP-----TSG 511
Qy 200 NOPDKLSTGAKLARDQPDLLTR--NOQCQFPVATQNTQFPMENQOAWLQMLNIGPFP 257
Db 512 LHPEDFIKKADLAVQKTPMINQGTQTEQNGQVMNITNSGHENKTGDSIQNE-----565
Qy 258 GNOQPRMTIRNOQPCPLAMGNQOPWYILGTPRALYSGNQOLGGPOGNKRPIFLN-HQTCL 316
Db 566 -----KNPNPIESLEKESAFKTKAAPTSSIS-NMELE-----LNTNSKA 605
Qy 317 PAGNOLYGSPTDMH-----QLVMS-----TGQOQHGLLIKNOQPGSLIRG 356
Db 606 PKNRLRRKSSSTRHIALELVVSRNLSPNCTELOIDSCSSSEEEKKKYNQMP---VRH 662
Qy 357 QQPCVPLIDQOPAT-----PKGFT-----HLNOMV 381
Db 663 SRNLQLMGKKEPATGAKKSNKPNQETSKRHSDTTPPELKLTNAPGSFTKCSNTSELKEFV 722
Qy 382 ATSM-----SSPGLRPHSQSOVP-TTYLHVES 407
Db 723 NPSLPREEKEKLETVKVSNNAEDPKDMLSGERVLOTERSSESSISLVPDGTGTQBS 782
Qy 408 VSRILNGTTGTQCR-----SRAPAYDSLQDDIH-----QGNKYILSHEISNGNG 451
Db 783 ISLLEVSTLGRKAKTEPNKCVSQCAAFENPKGLIHGCSKDNRNDTEGFKYPLGHEV-----837
Qy 452 CKKALPONSSLPPIWAKLEEA-----RQSKRQ-----YHRA 483
Db 838 -----NHSRETSI--EMEESLDAQYLQNTFKVSKRQSFALFSPNPGNAEECATFSAH 888
Qy 484 MGQTEKHDLNLAQOJTAQSDQDVERHNSSTCEVYLDAAKTKIKQVVQENL-HCMPPEVIEI 542
Db 889 SGLKKQSPKVTFECEKEENOGKNES-----NIRPQTVNITAGFP--VVGQ 934
Qy 543 EDDPTDGRAR--KGKNTASIS--KGASKGNSSPVKTAKEKCIVP-----KTPA 587
Db 935 KDKPVDNAKCSIKGSRFCLSSQFRNETGLITPNKHGLLQNPYRPLPIKPSFKTKC 994
Qy 588 KKGRAGRKSVPPPAHASEIQLWQTPPKT--PLRSRP-----KGKGRKSTODSGKARG 640
Db 995 KKNLL--BENFEHSMSPEREMGNIPSTVTSIRNNIRENVFKGASSNINEVGSSTN 1052
Qy 641 PSELLCQDSIAEIIYRMONLY--LGDKEREQOANWLYKGDGALVPVYESK-----691
Db 1053 EVG-----SSINEIGSSDENIOAELG--RNRGPKLAAMLRL---GVLPQEVYKQSLPGSNC 1103
Qy 692 RKP RPKVDIDDETTIRIWNL-----RADSFIAHMLVQDGRFRFPWKGSVVDSVI 721
Db 1104 KHPEIKKQYEVEVQTVNTDFSPYLLSDNLEQPMGSSHASQVCSFETPDLLDDGEIKEDT 1163
Qy 722 E--EKDKKKEKMWEEERRVF-----RG-----RADSFIAHMLVQDGRFRFPWKGSVVDSVI 771
Db 1164 SPAENDIK-----ESSAVFSKSVQSGELSRSPFTTHLAQGYRGA-----1206
Qy 772 GVFLTQNVSDHLSSSAFMSLAARFPFKLSSSRREDNRVSVVVEDEPE-GCINLNL---N 825
Db 1207 -----KKLESS---BENLSS-----EDELPCFOHLLFGKVN 1235
Qy 826 EIPSWQERKVOHPDMEVSGVDSGSKE-----QLRDCSNSGIERNFLEKSIQNLLEEV 878
Db 1236 NIPS--QSTRH-STVATECLSKNTEENLLSKNSLNDCSNQVI---LAKASQ---EHH 1284
Qy 879 LSSQDSFDPDAPQSCGRVSGSCSKSDAEFPPTTRETCTVTSQSVQTSQSPN-----LSD 934
Db 1285 LSEETKCSASLFS-----QCSELEDITANTQDPFLIGSSKQMRHQSESGVGLSD 1337
Qy 935 EICLOGNERPHLYEGSGDVQKQETTNVAKKPDLEKTNWKKDSVCFQGPDRNDTNQTPPS 994
Db 1338 RELVSDDEE---RGTG---LEENN--OEQOSMDNLEGAASGC-----ESET 1376
Qy 995 SSYEACATRQPHVLDIEDFGMOGEGLYSWNSISPRVDRVKNKNVPRRFRFGQGSVPREF 1054

Db 1377 SVSEDCS----- 1383
QY 1055 TGOIIPSTHELPGWGLSGSSSAVOEHQDDT-QHNQOENMKASHLOKTFDLNLSSEEC 1113
Db 1384 -----GLSSQSDILTTQORDTMQHN-----LIKLOQEMAELEAVLEQH 1421
QY 1114 LTRQSTKQNTDGCPLPRDRTAEDVVDPLNNSSLQNLVSNSSNKEQTAVEYKETNAT 1173
Db 1422 GSQPSNSVPSI-----ISDSSALEDLRNPQSTSEKAVLTSQKSSEYP 1464
QY 1174 ILREMKGTADGKKPTSQWDSLRLKDVCEGQRGRNKNMDSIDYEAIRRASISEISEAI 1233
Db 1465 ISQNPGLSAD-KFEVSADSTSKNKE--PCVERSPPSKCPSLD----DRWYMHSCSGSL 1517
QY 1234 KERGM-NNMLAVRIKDFLERIVKDHGGIDL 1262
Db 1518 QNRNYPQOEELIKVVDVEEQOLEESGPHDL 1547

Search completed: July 5, 2001, 12:43:30
Job time: 73 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 12:42:18 ; Search time 16.84 Seconds
(without alignments)
3517.086 Million cell updates/sec

Title: PCT-US01-13059-2

Perfect score: 9089

Sequence: 1 MQSIMSSAVNATEPQND.....PRPLMARLHPASKLKNKT 1729

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	246.5	2.7	3256	1 KI67_HUMAN	P46013 homo sapien
2	225.5	2.5	2375	1 ATRX_HUMAN	P46100 homo sapien
3	225	2.5	2468	1 MAPB_HUMAN	P46821 homo sapien
4	211.5	2.3	2842	1 APC_RAT	P70478 rattus norv
5	210.5	2.3	1805	1 NEST_RAT	P21263 rattus norv
6	208.5	2.3	3924	1 ANK2_HUMAN	Q01484 homo sapien
7	207	2.3	1723	1 AIM1_HUMAN	Q9V4K1 homo sapien
8	204.5	2.2	2717	1 ZEP1_HUMAN	P15822 homo sapien
9	200.5	2.2	3130	1 DPOZ_HUMAN	O60673 homo sapien
10	192.5	2.1	1781	1 AKAC_HUMAN	Q02952 homo sapien
11	192.5	2.1	1812	1 BRC1_MOUSE	P48754 mus musculus
12	192	2.1	2442	1 CBP_HUMAN	Q92793 homo sapien
13	191.5	2.1	2349	1 TPR_HUMAN	P12270 homo sapien
14	190	2.1	1284	1 ATL1_MOUSE	P16602 cowpox viru
15	190	2.1	2464	1 MAPB_MOUSE	P14873 mus musculus
16	189	2.1	2453	1 NCR1_MOUSE	Q60974 mus musculus
17	188.5	2.1	1435	1 EBA1_PLAFC	P19214 plasmodium
18	188.5	2.1	3396	1 PCGV_HUMAN	P13611 homo sapien
19	186.5	2.1	1132	1 YRK5_YEAST	P34250 saccharomyc
20	186.5	2.1	1833	1 ZEP2_HUMAN	P31629 homo sapien
21	186	2.0	1531	1 NTF5_HUMAN	O94916 homo sapien
22	184.5	2.0	1875	1 MLP1_YEAST	Q02455 saccharomyc
23	183	2.0	2441	1 CBP_MOUSE	P45481 mus musculus
24	182.5	2.0	1861	1 MAP2_RAT	P15146 rattus norv
25	182	2.0	2688	1 ZEP1_MOUSE	Q03172 mus musculus
26	181	2.0	1658	1 YMF7_YEAST	Q03661 saccharomyc
27	178	2.0	2230	1 GOG4_HUMAN	Q13439 homo sapien
28	178	2.0	2476	1 ATRX_MOUSE	Q61687 mus musculus
29	177.5	2.0	1972	1 P531_HUMAN	Q12888 homo sapien
30	177.5	2.0	2649	1 BPA1_HUMAN	Q03001 homo sapien
31	176.5	1.9	3358	1 PCGV_MOUSE	Q62059 mus musculus
32	175.5	1.9	2845	1 APC_MOUSE	Q61315 mus musculus
33	175	1.9	1210	1 AFA_HUMAN	P51825 homo sapien

ALIGNMENTS

RESULT 1

KI67_HUMAN
ID KI67_HUMAN STANDARD; PRT; 3256 AA.

AC P46013;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 01-OCT-2000 (Rel. 40, Last annotation update)

DE ANTIGEN KI-67.

GN MKI67.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=94043435; PubMed=8227122;

RA Schlueter C., Duchrow M., Wohlenberg C., Becker M.H.G., Key G.,

RA Flad H.-D., Gerdes J.;

RT "The cell proliferation-associated antigen of antibody Ki-67: a very

RT large, ubiquitous nuclear protein with numerous repeated elements,

RT representing a new kind of cell cycle-maintaining proteins.";

RL J. Cell Biol. 123:513-522(1993).

RN [2]

RP SEQUENCE OF 1-31 FROM N.A.

RA Gerdes J.;

RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: THOUGHT TO BE REQUIRED FOR MAINTAINING CELL

CC PROLIFERATION.

CC -!- SUBCELLULAR LOCATION: NUCLEAR. PREDOMINANTLY LOCALIZED IN THE G1

CC PHASE IN THE PERINUCLEAR REGION, IN THE LATER PHASES IT IS ALSO

CC DETECTED THROUGHOUT THE NUCLEAR INTERIOR, BEING PREDOMINANTLY

CC LOCALIZED IN THE NUCLEAR MATRIX. IN MITOSIS, IT IS PRESENT ON ALL

CC CHROMOSOMES.

CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A

CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -!- DEVELOPMENTAL STAGE: EXPRESSION OF THIS ANTIGEN OCCURS

CC PREFERENTIALLY DURING LATE G1, S, G2, AND M PHASES OF THE CELL

CC CYCLE, WHILE IN CELLS IN G0 PHASE THE ANTIGEN CANNOT BE DETECTED.

CC -!- SIMILARITY: CONTAINS 1 FHA DOMAIN.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; X65550; CAA46519.1; -

DR EMBL; X65551; CAA46520.1; -

DR EMBL; X94762; CAA64388.1; -

DR MIM; 176741; -

DR InterPro; IPR000253; -

DR Pfam; PF00498; FHA; 1.

DR PROSITE; PS50006; FHA_DOMAIN; 1.

KW Cell cycle; Antigen; Nuclear protein; ATP-binding; Repeat;

O14497 homo sapien
Q13061 homo sapien
P10388 triticum ae
P49454 homo sapien
P08799 dictyostell
Q09472 homo sapien
P55200 mus musculu
P25054 homo sapien
Q24742 drosophila
P38398 homo sapien
P08489 triticum ae
P15924 homo sapien

Db 3038 RCKSPVPVIMKRSRLR-----TSAKRIEPAEELNSNDM-KTNKEHKQLQDSVPENKIGISLR 3092

Qy 1532 LKNISRLRTEHQVVEL-PDSHRLLDGMDKREDPPSPYLLAIWTPGETA----- 1579

Db 3093 SRROKTEAEQOITEFVLAIEINRNEKKPMKTSPE-MDIQNPDGARKPIPRDKVTE 3151

Qy 1580 -----NSAQQP-EQKCGGKASGKCMFD-----ETCSECNLSLRANSQTVR 1618

Db 3152 NKRCLSARQNESSQPKVAEESGGKSAKVLQMNQKGEAGNSDSMLCRSKRTKSOPAA 3211

Qy 1619 GTL 1621

Db 3212 STL 3214

RESULT 2

ATTRX_HUMAN

ID ATTRX_HUMAN STANDARD; PRT; 2375 AA.

AC P46100; P51068; Q15886;

DT 01-NOV-1995 (Rel. 32, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE TRANSCRIPTIONAL REGULATOR ATRX (X-LINKED HELICASE II) (X-LINKED

DE NUCLEAR PROTEIN) (XNP).

CN ATRX OR RAD54L OR XR2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-97123494; PubMed=8968741;

RA Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.,

RA Gibbons R.J.;

RT "ATR-X encodes a novel member of the SNF2 family of proteins: mutations

RT point to a common mechanism underlying the ATR-X syndrome.";

RL Hum. Mol. Genet. 5:1899-1907(1996).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE-97386582; PubMed=9244431;

RA Villard L., Lossi A.M., Cardoso C., Proud V., Chiaroni P.,

RA Collea L., Schwartz C., Fontes M.;

RT "Determination of the genomic structure of the XNP/ATR-X gene encoding

RT a potential zinc finger helicase.";

RL Genomics 43:149-155(1997).

RN [3]

RP SEQUENCE OF 743-2375 FROM N.A.

RX MEDLINE=95179111; PubMed=7874112;

RA Stayton C.L., Dabovic B., Gullisano M., Gecz J., Broccoli V.,

RA Giovanazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,

RA Bianchi M.E., Consalez G.G.;

RT "Cloning and characterization of a new human Xq13 gene, encoding a

RT putative helicase.";

RL Hum. Mol. Genet. 3:1957-1964(1994).

RN [4]

RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.

RX MEDLINE=94214473; PubMed=8162050;

RA Gecz J., Pollard H., Consalez G., Villard L., Stayton C.,

RA Millasseau P., Khrestchatsky M., Fontes M.;

RT "Cloning and expression of the murine homologue of a putative human

RT X-linked nuclear protein gene closely linked to PK1 in Xq13.3.";

RL Hum. Mol. Genet. 3:39-44(1994).

RN [5]

RP SEQUENCE OF 2284-2375 FROM N.A., AND VARIANTS ATR-X.

RX MEDLINE=95211835; PubMed=7697714;

RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;

RT "Mutations in a putative global transcriptional regulator cause X-

RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";

RL Cell 80:837-845(1995).

RN [6]

RP VARIANTS ATR-X.

RX MEDLINE=97467722; PubMed=9326931;

RA Gibbons R.J., Bachoo S., Picketts D.J., Aftimos S., Asenbauer B.,

RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,

RA Levin M.L., Masuno M., Neri G., Pierpont M.E., Slaney S.F.,

RA Higgs D.R.;

RT "Mutations in transcriptional regulator ATRX establish the functional

RT significance of a PHD-like domain.";

RL Nat. Genet. 17:146-148(1997).

RN [7]

RP VARIANT JM GLN-2014.

RX MEDLINE=96224392; PubMed=8630485;

RA Villard L., Gecz J., Mattei J.-F., Fontes M., Saugier-Verber P.,

RA Munnich A., Lyonnet S.;

RT "XNP mutation in a large family with Juberg-Marsidi syndrome.";

RL Nat. Genet. 12:359-360(1996).

CC -!- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES

CC GENE EXPRESSION BY AFFECTING CHROMATIN. MAY BE INVOLVED IN

CC BRAIN DEVELOPMENT AND FACIAL MORPHOGENESIS.

CC -!- SUBCELLULAR LOCATION: NUCLEAR.

CC -!- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF X-LINKED ALPHA-

CC THALASSEMIA/MENTAL RETARDATION SYNDROME (ALSO KNOWN AS ATR-X

CC SYNDROME). ATR-X IS AN X-LINKED DISORDER COMPRISING SEVERE

CC PSYCHOMOTOR RETARDATION, CHARACTERISTIC FACIAL FEATURES, GENITAL

CC ABNORMALITIES, AND ALPHA-THALASSEMIA.

CC -!- DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF JUBERG-MARSIDI SYNDROME

CC (JM). JM IS A RARE X-LINKED RECESSIVE DISEASE CHARACTERIZED BY

CC SEVERE MENTAL RETARDATION, GROWTH FAILURE, SENSORINEURAL DEAFNESS,

CC MICROGENTALISM AND EARLY DEATH.

CC -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.

CC -!- SIMILARITY: CONTAINS 1 PHD-FINGER DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch)

CC -----

DR EMBL; U72938; AAB49971.1; -

DR EMBL; U72935; AAB40699.1; -

DR EMBL; U72904; AAB40699.1; JOINED.

DR EMBL; U72907; AAB40699.1; JOINED.

DR EMBL; U72908; AAB40699.1; JOINED.

DR EMBL; U72909; AAB40699.1; JOINED.

DR EMBL; U72910; AAB40699.1; JOINED.

DR EMBL; U72911; AAB40699.1; JOINED.

DR EMBL; U72912; AAB40699.1; JOINED.

DR EMBL; U72913; AAB40699.1; JOINED.

DR EMBL; U72914; AAB40699.1; JOINED.

DR EMBL; U72915; AAB40699.1; JOINED.

DR EMBL; U72916; AAB40699.1; JOINED.

DR EMBL; U72918; AAB40699.1; JOINED.

DR EMBL; U72919; AAB40699.1; JOINED.

DR EMBL; U72920; AAB40699.1; JOINED.

DR EMBL; U72921; AAB40699.1; JOINED.

DR EMBL; U72922; AAB40699.1; JOINED.

DR EMBL; U72923; AAB40699.1; JOINED.

DR EMBL; U72924; AAB40699.1; JOINED.

DR EMBL; U72925; AAB40699.1; JOINED.

DR EMBL; U72926; AAB40699.1; JOINED.

DR EMBL; U72927; AAB40699.1; JOINED.

DR EMBL; U72928; AAB40699.1; JOINED.

DR EMBL; U72929; AAB40699.1; JOINED.

DR EMBL; U72930; AAB40699.1; JOINED.

DR EMBL; U72931; AAB40699.1; JOINED.

DR EMBL; U72932; AAB40699.1; JOINED.

DR EMBL; U72933; AAB40699.1; JOINED.

DR EMBL; U72934; AAB40699.1; JOINED.

DR EMBL; U72936; AAB49969.1; -

DR EMBL; U72935; AAB40700.1; -

DR EMBL; U72908; AAB40700.1; JOINED.

DR EMBL; U72909; AAB40700.1; JOINED.

DR EMBL; U72910; AAB40700.1; JOINED.

DR EMBL; U72911; AAB40700.1; JOINED.

Db 1148 NDPENRIAKKML-----LEEIKANLSSD-----EDGSSDDPPECKKRTGKQ 1190
 QY 963 OKKPDLEKTMNKKVSCVCPQPNNDNWQTTSSSYEQCATRPHVLDIEDFGMGEGLY 1022
 Db 1191 EENPGDEAKQVNS-----ESDSSESKKPRYRHLR--HKLTVSD-GESE--- 1237
 QY 1023 SWMSISPRVDRVKNVPRFRQGSVPRETGOIIPSTHELPGMGLSGSSSAVQEHQ 1082
 Db 1238 -----EKKTKPEHKEVGRNRK-----VSESDSDFOESGVSEVSESE 1280
 QY 1083 DDTQHNQD-----EMNKASHLQKTFDLNLSSECLTROSSTKQNTDGLCPDRDAE 1136
 Db 1281 DQRPTRSAKKAEEENQYSKQ-----KRR-----RIKQVE 1315
 QY 1137 DVDPPLSNSSQNLTVSNNKQETAVEYKETWATILREMKGTLAGCKPTSDWSLR 1196
 Db 1316 D-----SSSEN---KSNSEEEEEEEEEEEEEDEEDNDSDSKSPGKRKKIR 1364
 QY 1197 KDVEGNEGRQERNKNMDSIDYEATRRASISE 1228
 Db 1365 KILKDDKLRTE-----TQNALKEEERKRRIAE 1392

RESULT 3

MAPB_HUMAN
 AC P46821; STANDARD; PRT; 2468 AA.
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE MICROTUBULE-ASSOCIATED PROTEIN 1B [CONTAINS: MAP1 LIGHT CHAIN LC1].
 GN MAP1B.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95104835; PubMed=7806212;
 RA Lien L.L., Feener C., Fischbach N., Kunkel L.M.;
 RT "Cloning of human microtubule-associated protein 1B and the
 RT identification of a related gene on chromosome 15.";
 RL Genomics 22:273-280(1994).
 CC -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
 CC PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
 CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
 CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
 CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
 CC STABILIZING MICROTUBULES.
 CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
 CC WITH MAP1A AND MAP1B PROTEINS.
 CC -!- DOMAIN: HAS A HIGHLY BASIC REGION WITH MANY COPIES OF THE SEQUENCE
 CC KKEE AND KKEI/V, REPEATED BUT NOT AT FIXED INTERVALS. THIS LATTER
 CC REGION IS RESPONSIBLE FOR THE BINDING OF MAP1B TO MICROTUBULES
 CC BOTH IN VITRO AND IN VIVO.
 CC -!- PTM: LC1 IS COEXRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED
 CC FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
 CC BOTH MAP1A AND MAP1B.
 CC -!- SIMILARITY: TO NEURAXIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; I06237; AAA18904.1; -;
 DR MIM; 157129; -;
 DR InterPro; IPR000102; -;
 DR Pfam; PF00414; MAP1B_NEURAXIN; 10.
 DR PROSITE; PS00230; MAP1B_NEURAXIN; 6.

KW Microtubules; Repeat; Phosphorylation.
 FT CHAIN ? 2468 MAP1 LIGHT CHAIN LC1.
 FT DOMAIN 589 790 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
 FT KKEE AND KKEI/V REPEATS).
 FT 12 X 17 AA TANDEM REPEATS.
 FT DOMAIN 1869 2074 1.
 FT REPEAT 1869 1885
 FT REPEAT 1886 1902 2.
 FT REPEAT 1903 1919 3.
 FT REPEAT 1920 1936 4.
 FT REPEAT 1937 1953 5.
 FT REPEAT 1954 1970 6.
 FT REPEAT 1971 1987 7.
 FT REPEAT 1988 2004 8.
 FT REPEAT 2005 2021 9.
 FT REPEAT 2022 2038 10.
 FT REPEAT 2039 2055 11.
 FT REPEAT 2056 2072 12.
 SQ SEQUENCE 2468 AA; 270618 MW; 540839CBDF09D461 CRC64;

Query Match 2.5%; Score 225; DB 1; Length 2468;

Best Local Similarity 18.5%; Pred. No. 0.00079;

Matches 336; Conservative 254; Mismatches 710; Indels 518; Gaps 81;

QY 175 LDSVSQIDKT---NGLGAMNOPLEVSNGNQDKLSTGAKLARDQDPLDLTRNQOCQRPV 230
 Db 301 LDRVDSILLTHIGDDNLPGLNSMLQRTAELEESQSGSTTNSDMWNKLLISPDLVGVFLN 360
 QY 231 ATQNTQFPMENQOAWLQMKNLQIGFPFNGQNPQPMTRINQOQCL-AMGN----- 277
 Db 361 VPENLKNPEPN---IKMRSIEEACFTLQYLNKLSMRPEPLFRSVGNITDIPVLFQKMG 416
 QY 278 --QQPMYLIIGTPPA--LVSGNQQLGG-----POGKRRIPLNHTQCL----- 316
 Db 417 VGKLEMYVLNPKSKSEKQYFQWGTGNTKKAETPLPNGQEVDPISPILTSVSSLLVWH 476
 QY 317 PAGNOLYGSPTDMHOLVMSTGQ-QHGLL-----IKN-----QOQSL---IRGQOQCVPL 363
 Db 477 PA-----NPAEKIIRVLFPNGNSTYNILEGLEKLUHLDLFLKOPLATQKDLGQVP-TPV 529
 QY 364 IDQOPATPKGFTHLNMQ-----VATSMSSPGLRPHSQSQVP--TTLHVESVSRIL 412
 Db 530 VKQ-----TKLQRADSRSLAPAAKPLPSKSVKRESKEETPEVTKVNHVE----- 575
 QY 413 NGTTGTCORSAPAYDSIQDDIHQGNKYILSHEISNGCKKALPQNSLSLPIPAKLEEE 472
 Db 576 -----KPKVESKEKVMVKDKPKVKE--TKPSVTEKEVPSKER-PSPVKAEEVAE 622
 QY 473 ARGSKRQYHRAMQTEKHDLNLAQOIAQSQ-----DVERHNSSTCVEYLDAAKTKIQ 526
 Db 623 KOATDVKPKAAKERTVKKTKVKPEDKKEEKPKKEVAKKEDKTPIRKKEEKPKKEEVKK 682
 QY 527 VVOENLHGMPPVEIETDDPTDGARKGNTASISGASKGNSPVKKTAEKKEKIVPKTP 586
 Db 683 EVKKEIKK-----EKKPKKEVAKKETPPKVEKVEKKEEKKVKK-EEKE----PKKE 731
 QY 587 AKKGAGRRKKSVPYPAHASEIQLWQTPP-----KTPLSRSKPKGKR-KSIQDSGKA 638
 Db 732 IKKLPKDAKKSSTPLSEAKKPAALPKVPKVEESVKDSVAAGKPKKGIKVIKKEGKA 791
 QY 639 RGPSELICQDSIAEIIYRMQNLVLDKEREQEQANVLYKGDGALVPYSEKRRPRPKV 698
 Db 792 AEAVAAAVGTGATTAATAVMAAAGIAAIGPAKELEAERSLMSSPEDLTQKDFEELKAE---EV 848
 QY 699 DIDDFTTRTNLMLGKDEKGEDEKDKKKEK-----WVEEERFRGRADSFIAARMHLVQ 754
 Db 849 DVTDKIKPQLELI-----EDEKLEKTEPEVYAVVQKEREVTKGPAESDEGITTTE 900
 QY 755 GD-----RRFSPMKGSVDSVIGVFLTQNVSHLSSSAFMSLAARFPFKPLSSREDER 807
 Db 901 GEGECEQTPPEELEPVKEQGVDDI-----EKFEDEGAGF-EESSETGDVEE 944
 QY 808 NVRSVVVEDDEG-----CILNLNIPSWQEK-----VOHPSDMEVSGVDSGSKEQ 852

Query Match 2.3%; Score 211.5; DB 1; Length 2842;
Best Local Similarity 18.4%; Pred. No. 0.0048;
Matches 319; Conservative 230; Mismatches 654; Indels 533; Gaps 83;

```
QY 64 AELPKVVE-----GPKRKRKAATOKVSKETGSAKKNLKESATKKPANV-----112
DB 883 AQIAKWEVVSALHTSQDSDSPASAEHLCHVAERTAA-----RSSASHPTNHNFAK 937
QY 113 GDMNKSPEVTL-----KSCRKALNFLENPCDARQDSESEIVQNSGANSFSEIRDA 166
DB 938 SESSNRCTMPYAKVEYKRSSDLSNVTSSDGYGKRG-----QMKPSVESYEDDSG 990
QY 167 IGTNGSP-LDSVSIQIDKTNGL-----GANNQPLEVSMGNQPKLSTG-----AKLARD 214
DB 991 KFCYSQGVYADLAHIXHSANHHDDNGGELDTPTINYSKYSDQLNSGROSPQSNERWARP 1050
QY 215 QO--PDLLTRNOQCOFFPVATQTPMENQOAWLQMKNLQIGFPFGNQOQPRMIRNOQPC 272
DB 1051 KHVIEDEIKONEQRO--SRSQNTNFPVISENT--DDKHLKFOQHFG-----QQEC 1096
QY 273 LAMGNOQPMYLLIGTPRALVSGNOQLGGPQGNKRPIFLNHQTC-----LPAGNOLYG 324
DB 1097 VS-----PYRSRGNG-----SETNRMGSSHAVNQNV--NQSLCQEDDYEDDKPTNYSERY 1145
QY 325 SPTDMHQ-----LWMTGGGOHGLLKNQO-----GSLIRGOQPCVPLIDQOQA 369
DB 1146 SEEOHEEERPTNYSIKYNEEKHHV-----DQPIDYSLKYATDISSQKPSFS-FSKTSP 1200
QY 370 TPKGTHLNQWATSMSPGLPHSPQSQVPTTYLHVESVSRILNGTTGTCQSRAPAYDS 429
DB 1201 VOGTTEHNSPSEASAPSSNAKRQSO-----LHPSSAQR--NGQTPKGTACKVP--S 1250
QY 430 LQODIHQGNKYILS-----HEISNGN--GCKKALPQNSLPTPIMAKLEEAR 474
DB 1251 INQETMQ--TYCVEDPTICFSCSSLSLSAEDELGCDDTQOEDASANTLOIAEIKEND 1308
QY 475 GSKROYH-----RANGOTEKHLNLAQQAQODVERHNSCTCYLDAK---KTKIQK 526
DB 1309 VTRSAQDPASDPVAVSQSTRTPSRLOASGLASEARHKA---VEFSSGAKSPSKSGAQT 1365
QY 527 VVOENLHGMPPVEIEDDPTDGARKGKTATASIKGASKNSPVKKTAEKECI-----581
DB 1366 PKS-----PPHY--VOETPLVFSR-----CTSVS--SLDSFESRSIASSVQSEPCSGMVG 1413
QY 582 -----VPKTPAKGRAGRKSVPPPAHASEIQLMQPTPKTPLSRKPGRKRSIQDS 635
DB 1414 IVSPSDLPSQGTWPPPSKSTPPP-----PPQPVQTKREVPTKVPAAEQRE 1463
QY 636 GKARGPSGELLQDSIAETIYRMQNLVLDGK-----EREQEQNAMVLYKGDGAL---VPYE 688
DB 1464 G---GPK-----QTAVSAAVQVQVLPDADTLHLHFATESTPDGFCSSLSALSLEDEPI 1515
QY 689 SKRKRPR--PKVDIDDETTRINLLMGKDE-----KEGDEKDKKKKEKWEHEE-----735
DB 1516 QKDVELRIMPVOEND-----NGNETEPEQPEESNQDKEVEKPDSEKDLDD 1564
QY 736 -----RRVER-----GRADSFIAARMHLVQGDNR---758
DB 1565 SDDDDIEILEECIISAMPTKSRKAKKLAQTASKLPPVPARPSQLPVYKLLPSQSRLOA 1624
QY 759 -----FSPWKGSVDSVIGVFTQNVSDHLSSAFMS--LAARFPKLSRREDERNRVS 812
DB 1625 OKHVSFTPG-----DDVPRVYCEGTPINFSTATSLDSTIESPPNELAAGD---GVRAS 1676
QY 813 VVEDPEGCTILNLETSPWQEKVQHVSMDMEVSGVD-----SGSKEQ-----LRDCNSGIERF 864
DB 1677 V---OSGEFEKRDITPTTEGRSTDEAQRGVSSIAIPDLDGSKAEEDGILAEICNSA---1729
QY 865 NFLEKSIONLEEVLSQDSFDPATFQSG-----RV 896
DB 1730 --LPGRSHKPPRVKIMDQVOQAQSWTSSGTTNKNQIDTKKKKPTSPVKMPONTEYRTRV 1787
```

```
QY 897 GSCSCSKSDAEPTTRCETKTVSGTSQSQVOTGSPNLSDEICLOGNE-----RPHL 946
DB 1788 RNTDSKVNUNVTEETFSDNK--DSKQKSLKNNPKDLNDR--LPDNEDRVRGGTFDPSPHH 1843
QY 947 Y---EG-----SGDVQKQETTINVAOKKPDLEKTMNWKDSVCFQCPQRNDTNWOT 991
DB 1844 YAPIEGTPYCFSRNDSLSLDFD-DDVDVLSREKAELRGKESKDS-----EAKVYCHT 1896
QY 992 TPSSSYE-----OCATROPHVLDIEDFGMOGEGGLGYSWMSISPRVDRVKNKNVPRRFFRQ 1046
DB 1897 EFSSSQOASARKAAQASTKHP-----VNRGPSKPLLEQEQTPFP--SKKDVPDR---1941
QY 1047 GGSVPREFTGIIIPSTPHELPGMGLSGSSSAQVEHODDTHNOODEMNKASHLOKTFELD 1106
DB 1942 GAATDEKLFQNFALNETP-----VCFSRNSSLSDSDVOEQNNNEETGPV-----1986
QY 1107 LNSSECLTRQSTQKITDGLCPDRDAEDVVDPLSNSSNQILVESNNSNKEQTAVE 1166
DB 1987 --RDAEPANAQOPGKPOASGYAPKSFHYEDTPVCFSRNSSLSDSIDSEDD-----2036
QY 1167 YKETNATILRE-MKGTLDAGKKPT-----SQWDSLRL-----KDVEGNEGRQ 1206
DB 2037 -----LLRECISAMPKKRRPSRLKGEQWQSPRKVGSVLAEDLTLDLKDQRPSEH 2089
QY 1207 ERKNKMSIDYEATRASISEISEAIEKRGMMNMLAVRIKDFLERIVKDHGIDLEWLR 1266
DB 2090 GLSPDS--ENFDKATQOEGANSIVSSLHQAASAAASQASDSDSILSKGVSIGSPF 2148
QY 1267 ESPPOKADYLLSIRGLGLKSVCEVRLTLHLNLAFPVDTNVGRIAVRMGWVPLQPLPESL 1326
DB 2149 HLTPOEEKPTFSHGK-----2164
QY 1327 QLHLELYPVLESIOKFLWPRCLKDQRTLYELH-----YOLITFGKVFC 1371
DB 2165 -----PRILKPEKSTLEAKKTESENKGIKGGKVKYKSLITSKI--2203
QY 1372 TKSRRNCN--ACPMRGECRHFASAYASARLALPAPERSTSATIPVPPSPFPVPAIPMIE 1430
DB 2204 ---RSNSEISSOMKOPLOTNMPISIRGMTIHIPGVNRNSSSTSPVSKRG-PP-----2252
QY 1431 LPLPLEKSLASG--APSNRENCEPIIEEPASQOECTETESDIEDAYYNEDP-----D 1482
DB 2253 LKTPASKSPSEGVATTSPRGTKPAVKSELSP-----ITROTSHISGSKNGPSRSGSRD 2306
QY 1483 EIPTIKLNIQEGMTLREHMERNMELQEGDMSKALVALHPTTTTSIPT-CLKNISR 1537
DB 2307 STPS-----RPTQOPLSRPMQ-----SPGRNSISPGRNGISTPNKLSQLPR 2347

RESULT 5
NEST_RAT
ID NEST_RAT STANDARD; PRT; 1805 AA.
AC P21263;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE NESTIN.
OS NES.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90150286; PubMed=1689217;
RA Lendahl U., Zimmerman L.B., McKay R.D.G.;
RT "CNS stem cells express a new class of intermediate filament
protein."
RL Cell 60:585-595(1990).
CC -!- TISSUE SPECIFICITY: CNS STEM CELLS.
CC -!- DEVELOPMENTAL STAGE: UPON TERMINAL NEURAL DIFFERENTIATION, NESTIN
IS DOWN-REGULATED AND REPLACED BY NEUROFILAMENTS.
CC -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
```

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M34384; AAA41685.1; -
 CC PIR: A34736; A34736.
 CC InterPro: IPR001664; -
 CC Pfam: PF00038; Filament; 2.
 CC PROSITE: PS00226; IF; 1.
 CC Intermediate filament; Coiled coil; Heptad repeat pattern;
 CC Neurone.
 CC
 CC DOMAIN 1 7 HEAD.
 CC FT 8 314 ROD.
 CC FT 315 1805 TAIL.
 CC FT 43 43 COIL 1A.
 CC FT 44 55 LINKER 1.
 CC FT 56 151 COIL 1B.
 CC FT 152 174 LINKER 12.
 CC FT 175 193 COIL 2A.
 CC FT 194 196 LINKER 2.
 CC FT 197 314 COIL 2A.
 CC SEQUENCE 1805 AA; 198744 MW; B40EE14717E0998D CRC64;

Query Match 2.38; Score 210.5; DB 1; Length 1805;
 Best Local Similarity 18.78; Pred. No. 0.0029;
 Matches 283; Conservative 216; Mismatches 564; Indels 447; Gaps 72;

Qy 60 PRKPAELPKVVEGKPKRKAATQEVKSK--ETGSAKKKLNKESATKRPANVGDMNS 117
 Db 159 PRP-----PAPPHRIPGAPVEDELLARRLGEVVRGAVRDYQERVAAHMESSLGQARE 210
 Qy 118 KSEVTT--LKSCRKALNFDLENPGDARGDSEIVQNSSGANSFSEIRDAI--GGTNGSF 174
 Db 211 RLSQAVRGARECR-----LEVQQLQADRDLSLQERREALQRLGSRW 251
 Qy 175 LDSVSDIKTN--GLGAMNQPLEVSMGNQPKLSTGAKLARDQDPLLRNQOQFPVATQ 233
 Db 252 QDRLOATDKFQAVLAEQERKQGLSQTAQLGGQQLAH-----LKMSLSEVATY 303
 Qy 234 NTQPFMENOQAWLQKMLGIFPGNQOPRMTIRNQOPCLAMNQOPMYLIGTPPALVS 293
 Db 304 RTLEAENS--LOTGPGSGASLGLFDPKL-----KPNF--LGIPEDQYL--GSVLPALSP 354
 Qy 294 GNQQLGGPQGNKRPI--FLNHQTCIPAGNQLYSGPTDMHQLVMSTGGQHQGLLIKNOQPG 351
 Db 355 TSFSPPLNTLETPTVATFLKTQEFLOARTPTLAS--TPIPPISAP-----CPN 402
 Qy 352 SLIRGQPCVPLIDQOATP--KGTHLNQWATSMSSPG-----LR 391
 Db 403 AEVRAQEVPLSLQTOAPEPLWATVPSSAILPELEPGKQKQGHFPDDLTLATNLN 462
 Qy 392 PHSOSQVPTVTLHVESVRLNGTTCQSRAPAYDSLOQDIHQGNKI-----LSHEI 446
 Db 463 PHH-----PT-----LEAKDGSSSRVSSI--FOEDGOIWEIVKEADIEVKV 505
 Qy 447 SNGNCKKALPQNSLPTPIINAKLEEARG--KROVHRAMGQTEKHDLNLAQOIAQSQ 502
 Db 506 ENSSAQK--TQESGLDT-----EEQDSQGPLOKETLKAALG--EPLMSLKIONYETA 554
 Qy 503 DVERHNST-----CVEYDRAKTKTKQVVQENL-----HGMV--PEVIEDDPTD 548
 Db 555 GKNCNSSTEGHGTLEGPKEKQIPLSLKEKNVSEKTLENGVPVLSSELLGKDRTE 614
 Qy 549 -----GARK-----GKNTASISKAGSKN-----SSPVKKT 574
 Db 615 DQELMSPKGLTKRFSLSLGKESQEVVVRFSKEGNLESWTAFAKESOHPLGFGAEDQMLERL 674

Qy 575 AEKEKICVPKTPAKKG-----RAGRKKSVPPPAH-----ASEIQLMQPTPPK 616
 Db 675 VEKEDQSFPRSPEDQACPLQKNEQPLGVEAEQILERLIEKESQESLASPEED 734
 Qy 617 TPLSRSPKPKG---GRKSIDQSKARGPSELGLCDQSIABIIYRMQNLYLIDGR--ERQOE 671
 Db 735 QEAGRSQKQENQPLGVEAEADQMLER-----LIEKESQESLKSPEENQRIKPLERENQ 789
 Qy 672 QNAMVLKGDGALPYESKRRKPRKVDIDDITRIWNLMLGKDEKGD--EEDKKKEK 730
 Db 790 KSLYLEENQETTFPLESRNORPLRSLEVEEEOIRIVKPLEKVSQDSGLSLAENQVQLR 849
 Qy 731 WVEEE-----RRVFRGRADSFIAHMLVQGDRRFPKWSGVVDSVIGVFL-----TQNV 780
 Db 850 YLEEDDCINKSLLEDKTHKSLGSLDRNGDSIIIP-----QESETQVSLRPEEEDORIV 904
 Qy 781 DHLSSAFMSLAARPPKLSRSSREDERNV--RSVVVEDPEGCIINLNEIPWQKQVQHP 838
 Db 905 NMLEKES-----QFSSRSSEEEQVMERSLEGENHE---SLSSVEKEDQME--S 949
 Qy 839 DMEVSGVDG-----SKEOLRDCNSGIERFNFLKSIONL-----EE 876
 Db 950 QLEKESQDSGKSLSEDESOETFGPLEKNAESLRSLAGDQDEEQLEQETQOTFLAVGNEQ 1009
 Qy 877 EVLSSQDSFDP-----AIFQSCGRVSGCS--CKSDAEFFPTR--CETKTVSGTS-- 922
 Db 1010 MAVSPKPKVDPLKPLGNDQEIARSLGKENQESLVSLKEKGIETVKSLETEIIEPLETA 1069
 Qy 923 -----QSVOTGSPNLSDIEICLOGNERPHLYE---GSGDVQKQETNVAKKPDLEKT 971
 Db 1070 EEDLRKSIDTQPLMSTEVARETVEPPDEPPGSLGSDVDENRETTLSLEKESQESL 1129
 Qy 972 MNW-----KDSVCGFQPRNDTN-----WQTPPSSSYEQ----- 999
 Db 1130 GKWNVETVEDSQOCLQVEEGLOEQHESLRVQKQELPSSGNQORWEDVVEGKAVGOEA 1189
 Qy 1000 -CATRQPHVLDIEDFGMOGEG-----LGYSMW--SISPRVDRVKNKNVPR 1041
 Db 1190 PLATTGVGTEDKAEHLRGGGEEAAAGELLQDIVGEAWSLGSSSEPKQORVPAEALDN 1249
 Qy 1042 RFFQSGSVPREFTGQIIPTPHELPGMLGSLSSSAVOEHODD--TOHNOQDEMKNASHLQ 1100
 Db 1250 ---LEGGAL-----EVP---VAQSMPEVTERDEDAQAGEQDSIEVTGL 1289
 Qy 1101 --KTFDLNLSSECLTRQSSSTKQNTDGLPRDRTAEVDVPLSNSSLNQILVESNSS 1158
 Db 1290 AARTGLEL-----EQEVVGLDPRHFAEEAIPPSIGSESVK----- 1326
 Qy 1159 NKEQTAVEYKETNATILREMKGTIADGKKPTSQNDLSLRKQVEG--NEGQERKNKNMDSID 1217
 Db 1363 CQGHEE---SESMEGWEEEAAS-----LE--TSDHEGSDAQPFR--PPTEDEEG 1405
 Qy 1276 YLLSIRGLGLKSVQVRLTLHLNLAFPVDTVNGVRIARVMGWVLOPLPESLQLHLELYP 1335
 Db 1406 AQAALTAPGPKLLE-----PCSPILPILTAH--ELQP 1435
 Qy 1336 VLESIQKFLW 1345
 Db 1436 QAEGIQEAGW 1445
 RESULT 6
 ANK2_HUMAN
 ID ANK2_HUMAN STANDARD; PRT; 3924 AA.
 AC Q01484; Q01485;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE ANKYRIN 2 (BRAIN ANKYRIN) (ANKYRIN B) (ANKYRIN, NONERYTHROID).

GN ANK2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain stem;
RX MEDLINE=94075409; PubMed=8253844;
RA Chan W., Kordeli E., Bennett V.;
RT "440-kD ankyrinB: structure of the major developmentally regulated
RT domain and selective localization in unmyelinated axons.";
RL J. Cell Biol. 123:1463-1473(1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC TISSUE=Brain stem;
RX MEDLINE=91302466; PubMed=1830053;
RA Otto E., Kunimoto M., McLaughlin T., Bennett V.;
RT "Isolation and characterization of cDNAs encoding human brain
RT ankyrins reveal a family of alternatively spliced genes.";
RL J. Cell Biol. 114:241-253(1991).
RN [3]
RP REVISIONS.
RA Carpenter S.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 463-495 FROM N.A.
RX MEDLINE=92009921; PubMed=1833308;
RA Tse W.T., Menninger J.C., Yang-Feng T.L., Francke U., Sahr K.E.,
RA Lux S.E., Ward D.C., Forget B.G.;
RT "Isolation and chromosomal localization of a novel nonerythroid
RT ankyrin gene.";
RL Genomics 10:858-866(1991).
CC -!- FUNCTION: ANKYRINS ATTACH INTEGRAL MEMBRANE PROTEINS TO CYTO-
CC SKELETAL ELEMENTS; THEY BIND TO THE ERYTHROCYTE MEMBRANE PROTEIN
CC BAND 4.2. TO NA-K ATPASE. TO THE LYMPHOCYTE MEMBRANE. PROTEIN GP85,
CC AND TO THE CYTOSKELETAL PROTEINS FODRIN, TUBULIN, VIMENTIN AND
CC DESMIN. ERYTHROCYTE ANKYRINS ALSO LINK SPECTRIN (BETA CHAIN) TO
CC THE CYTOPLASMIC DOMAIN OF THE ERYTHROCYTES ANION EXCHANGE PROTEIN;
CC THEY RETAIN MOST OR ALL OF THESE BINDING FUNCTIONS.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL
CC CELLS THROUGHOUT THE BRAIN.
CC -!- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES
CC AND EACH PHOSPHORYLATION EVENT REGULATES THE PROTEIN'S STRUCTURE
CC AND FUNCTION (POTENTIAL).
CC -!- SIMILARITY: CONTAINS 23 ANK REPEATS.
CC -!- SIMILARITY: CONTAINS 1 DEATH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z26634; CAB42644.1; -;
DR EMBL; X56957; CAA40278.1; -;
DR EMBL; X56958; CAA40279.2; -;
DR EMBL; M37123; AAA62828.1; -;
DR PIR; S14533; S14533.
DR PIR; A39643; A39643.
DR PIR; B39643; B39643.
DR PIR; S14569; S14569.
DR HSP; Q00420; IAWC.
DR MIM; 106410; -;
DR InterPro; IPR000488; -;
DR InterPro; IPR000906; -;
DR InterPro; IPR002110; -;
DR Pfam; PF00791; ZU5; 1.
DR Pfam; PF00023; ank; 22.

DR Pfam; PF00531; death; 1.
DR PROSITE; PS50088; ANK_REPEAT; 20.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS50017; DEATH_DOMAIN; 1.
KW Cytoskeleton; Alternative splicing; Repeat; ANK repeat;
KW Phosphorylation; Multigene family.
FT REPEAT 63 92 ANK 1.
FT REPEAT 96 125 ANK 2.
FT REPEAT 129 158 ANK 3.
FT REPEAT 162 191 ANK 4.
FT REPEAT 193 220 ANK 5.
FT REPEAT 232 261 ANK 6.
FT REPEAT 265 294 ANK 7.
FT REPEAT 298 327 ANK 8.
FT REPEAT 331 360 ANK 9.
FT REPEAT 364 393 ANK 10.
FT REPEAT 397 426 ANK 11.
FT REPEAT 430 459 ANK 12.
FT REPEAT 463 492 ANK 13.
FT REPEAT 496 525 ANK 14.
FT REPEAT 529 558 ANK 15.
FT REPEAT 562 591 ANK 16.
FT REPEAT 595 624 ANK 17.
FT REPEAT 628 657 ANK 18.
FT REPEAT 661 690 ANK 19.
FT REPEAT 694 723 ANK 20.
FT REPEAT 727 756 ANK 21.
FT REPEAT 760 789 ANK 22.
FT REPEAT 793 822 ANK 23.
FT DOMAIN 1773 1950 REPEAT-RICH REGION.
FT REPEAT 1773 1784 REPEAT A.
FT REPEAT 1785 1796 REPEAT A.
FT REPEAT 1797 1808 REPEAT A.
FT REPEAT 1809 1820 REPEAT A.
FT REPEAT 1821 1832 REPEAT A.
FT REPEAT 1833 1844 REPEAT A.
FT REPEAT 1845 1856 REPEAT A.
FT REPEAT 1857 1867 REPEAT A.
FT REPEAT 1868 1879 REPEAT A.
FT REPEAT 1880 1891 REPEAT A.
FT REPEAT 1892 1902 REPEAT A.
FT REPEAT 1903 1914 REPEAT A.
FT REPEAT 1915 1926 REPEAT A.
FT REPEAT 1927 1938 REPEAT A.
FT REPEAT 1939 1950 REPEAT A.
FT DOMAIN 3536 3620 DEATH DOMAIN.
FT VARSPLIC 1039 1039 Q -> QFLGKLHLPTAPPLNEGESILVSRILQGPPTGK
FT (IN ISOFORM 2).
FT VARSPLIC 1444 3528 MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT CONFLICT 475 476 GQ -> PE (IN REF. 4).
FT CONFLICT 971 971 I -> S (IN REF. 2).
FT CONFLICT 3581 3582 QY -> HA (IN REF. 2).
FT CONFLICT 3586 3586 I -> Y (IN REF. 2).
SQ SEQUENCE 3924 AA; 430337 MW; 52AC496C428E29D2 CRC64;

Query Match 2.3%; Score 208.5; DB 1; Length 3924;
Best Local Similarity 18.0%; Pred. No. 0.011;
Matches 338; Conservative 235; Mismatches 676; Indels 625; Gaps 87;

QY 70 VVEGPKRPRKRAQTEQVKYKSTGSAKKNLKESATKKPANVGDMNSPEVTLKSCRK 129
DB 2185 LMEGTP-----QISSEE--SYKHEGLAETPTSPESLSFPKKSEQT----- 2225
QY 130 ALNFDLEPNGDARQGDSESEIVQNSSGANSFSEIRDAIGT--NGSFL--DSVSQIDKTN 185
DB 2226 -----GETKE-STKTETTTETSEKHEHTTKDITGGSERGATVTEDETSTESFQ 2275
QY 186 GLGAMNQPLEVSGMGNQPKLSTGAKLARDQQPDLLTRNQCCQ-----FPVATONTQFPME 240
DB 2276 KEATLGSPPKDTSPKQDDCTGSCSVALAKETPTGLTEEAACDEGQRTFGSSAHKQTQDSE 2335
QY 241 NQQAQ-----LQMKNLIGFPFGNQ-QPRWTRNQPC-LAMGNQOPMYLIGTPRA 290

```

Db 2336 AQESTATSDTKALPLPPEASVKTDTGTESKPGQVIRSPQGLEALPDSSEVL SAVADDS 2395
QY 291 LVSGNOQLGGPOGNRPFLFHNQHTCLPAG-----NOLYGSPTDMHQLVMSGGQHQGLLI 345
Db 2396 LA-----VSHKSLASPLVDENSHKTPDLSL-----2422
QY 346 KNOQPGSLIRGQPCVPLIDQOPATPKGFTHLNQVATSMSSPGLRPHRSQSQVPTTLHV 405
Db 2423 ---EPSPL--RESPCROSLSSPVEPK-----MKAGIFP-SHPPLPAVAKT 2463
QY 406 ESV-----SRILNGTTGTCQRAPAYDSLOD--IHQGNKYLSHEI--SNGNGCKKA 455
Db 2464 ELLTEVASVRKLLRDPDGSAD-----DSLEQTSMESSGKSPSPDTPSSSEVSVEV 2517
QY 456 LPONSSLPPTIMAKLEA-----RGSKROYHRAMGOTEKHDL-----NLAQO 497
Db 2518 TPKTTDVSTPAVHIECAEEDDSENGEKKRP-----TPEEMFKVWTKMKMFEDELEQE 2571
QY 498 IAQSDQV-----BRHNSST-----CVELDAAKTKTKQVQVQENLHGMPEVIEIEDDP 546
Db 2572 AKQKDYKKEPKQEBESSSSDPDADCSVDVDEPKHTG-----SGEDSGVPLVTSERKV 2627
QY 547 TDGARKGNTASISKGASKG-----NNSPVKTAKEKCIYVKP--584
Db 2628 SSSSESEPELAQLKKGADSGLLPEFVIRVQPPSPPLPSMDSNSPEEVQFQP-----VVSQK 2684
QY 585 -----TPAKKGRAGRKKS-----VPPAHASEIQWLQWTPPKTPLRSRKP-----KGKR 629
Db 2685 YTFKMNEDQEPKSEKESHESHLADRHAVSTEAEDRSYDKLNRDTPQPKICDGHGC 2744
QY 630 KSIQDSGAR-----GPSGELLQDSIAE--IYIRMONLYLGDKEREQONAMWLYKG 680
Db 2745 EAMSPSSSARPVSSGLOSPGTG-----DDVDEQPVYIKESLALQGTHEKDTGEEL-----2794
QY 681 DGALVPYESKRRKRPKVDIIDDE-----TTRIWNLLMGKDGKDEG-----721
Db 2795 -----DVSRAESPOADCPSESFSSSLPHCLVSEKDELDEISATSIQKTEVTKT 2846
QY 722 EEKDKKKEKWEERVRGRADF--IARHLVQGDRRSPKGSVVD--769
Db 2847 DETFENLPKQCPQSDSITTTQDRFSDMDVPVSDLAENDIYDQITSPYENVPSQSFSS 2906
QY 770 -----VIGVELTQNVSDHL-----SSSAFMSLAARF-----795
Db 2907 EESKTQTDANHTTSHSSEYVSVTITSPEDVVVASSSGTIVLSKESNFEQDIKMSQL 2966
QY 796 -----PPLKSSRDERNVRSVV--EDPEGILNLNEI--PSWQE-----KV 834
Db 2967 ESTLWEMQSDSVSSFEPTSATTVVGEQISKVITKTDVDSDSWSEIREDDAEFAEV 3026
QY 835 QHPSDMEVSG--VDSGSK-----EQ-----LRDCNSNGI--861
Db 3027 KE-BEQKIFGLMVRQSQGTTPTDTPARTPEETPTSEQNPFLOFGKLFEMTRSGAID 3085
QY 862 -----ERENFL-----EKSIONLEEVLSSQDSFPAIPQSGRVGSCSKSDA---906
Db 3086 MTKRSYADESPHFQIGQESREETLSEDDYKBGATGADPLPLETSAESIALSLESKETVDDE 3145
QY 907 -----EFPYTRCETKIVSGTSQ-----VOTGS--PNLS--933
Db 3146 ADLLPDSVSEVEEIPASDAQNLQMSQGISASTETPTKEAVSVGKDLPTVOTGDIPLPSG 3205
QY 934 -DEICLOGNERP-----HLYEGSGDVQKQETTNVAOKKPDLEKTMNWKDVC 979
Db 3206 VKQISCDSPSEPAVOQLDFSTLRSVYSDRGD--DSFDSSEPEQKSVIELPTAPMENVP 3263
QY 980 FGQPRN-----DTNWOTTPSSSYEQACATRQPHVLDIEDFGMOGELGYSWMSISPRVD 1032
Db 3264 FTESKIKIPVTMTPTSTPAPSAEYSSVS-----EDF-----LSSVDEENKAD 3307
QY 1033 RVKKN-----VPRFRFRQ-----GGSVPREFT--GQIIPSTPHELPGMLSGSSSAQVE 1080

```

```

RESULT 7
AIM1_HUMAN STANDARD; PRT; 1723 AA.
ID Q314K1; Q00296;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ABSENT IN MELANOMA 1 PROTEIN.
GN AIM1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97250519; PubMed=9096375;
RA Ray M.E., Wistow G., Su Y.A., Meitner P.S., Trent J.M.;
RT "AIM1, a novel non-lens member of the betagamma-crystallin
superfamily, is associated with the control of tumorigenicity in human
malignant melanoma."
RL Proc Natl Acad Sci U S A. 94:3229-3234(1997).
CC -I- FUNCTION: MAY FUNCTION AS SUPPRESSOR OF MALIGNANT MELANOMA. IT MAY
CC EXERT ITS EFFECTS THROUGH INTERACTIONS WITH THE CYTOSKELETON.
CC -I- SIMILARITY: BELONGS TO THE BETA/GAMMA-CRYSTALLIN FAMILY. CONTAINS
CC 6 TANDEM REPEATS OF A BETA/GAMMA-TYPE CRYSTALLIN DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

[illegible]

RA Omichinski J.G., Ciore G.M., Robien M., Sakaguchi K., Appella E.,
RA Gronenborn A.M.;
RT "High-resolution solution structure of the double Cys2His2 zinc
RT finger from the human enhancer binding protein MBP-1.";
RL Biochemistry 31:3907-3917(1992).
CC -1- FUNCTION: THIS PROTEIN SPECIFICALLY BINDS TO THE DNA SEQUENCE
CC 5'-GGGACTTCC-3' WHICH IS FOUND IN THE ENHANCER ELEMENTS OF
CC NUMEROUS VIRAL PROMOTERS SUCH AS THOSE OF SV40, CMV, OR HIV1.
CC IN ADDITION, RELATED SEQUENCES ARE FOUND IN THE ENHANCER ELEMENTS
CC OF A NUMBER OF CELLULAR PROMOTERS, INCLUDING THOSE OF THE CLASS I
CC MHC, INTERLEUKIN-2 RECEPTOR, AND INTERFERON-BETA GENES. IT MAY ACT
CC IN T-CELL ACTIVATION.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- INDUCTION: BY MITOGEN AND PHORBOL ESTER.
CC -1- DOMAIN: CONTAINS TWO SETS OF 2 ZINC-FINGERS, WHICH ARE WIDELY
CC SEPARATED AND RECOGNIZE THE SAME DNA SEQUENCE. THERE IS A FIFTH
CC ZINC-FINGER IN-BETWEEN.
CC -1- SIMILARITY: STRONG, TO HIVP2.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X51435; CAA35798.1; -;
DR PIR: A34203; A34203.
DR PDB: 3ZNF; 15-JAN-92.
DR PDB: 4ZNF; 15-JAN-92.
DR PDB: 1BBO; 31-OCT-93.
DR TRANSFAC: T00497; -;
DR MIM: 194540; -;
DR InterPro: IPR000822; -;
DR Pfam: PF00096; zf-C2H2; 5.
DR PRINTS: PR00048; ZINC-FINGER.
DR PROSITE: PS00028; ZINC FINGER C2H2.1; 4.
DR PROSITE: PS0157; ZINC FINGER C2H2.2; 4.
KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW Nuclear protein; Repeat; 3D-structure.
FT DOMAIN 406 456 ZINC FINGERS.
FT ZN_FING 406 428 C2H2-TYPE.
FT ZN_FING 434 456 C2H2-TYPE.
FT DOMAIN 803 806 POLY-SER.
FT ZN_FING 958 981 C2HC-TYPE (POTENTIAL).
FT DOMAIN 2087 2139 ZINC FINGERS.
FT ZN_FING 2087 2109 C2H2-TYPE.
FT ZN_FING 2115 2139 C2H2-TYPE.
FT STRAND 2088 2088
FT TURN 2090 2092
FT TURN 2095 2095
FT STRAND 2099 2108
FT HELIX 2099 2108
FT TURN 2109 2109
FT STRAND 2115 2116
FT STRAND 2123 2124
FT STRAND 2127 2135
SQ SEQUENCE 2717 AA; 297217 MW; D45D3CA951FEA561 CRC64;

Query Match 2.2%; Score 204.5; DB 1; Length 2717;
Best Local Similarity 16.9%; Pred. No. 0.01;
Matches 355; Conservative 297; Mismatches 706; Indels 741; Gaps 93;

QY 14 EATQNDG---SRQDLFEFLNKTPOQ-KPSKRKRKPKVYVGGKPKRKPAPLKV 69
DB 18 EAQELNGAEVSKREILQAGVKGTSES�GKVRK-----KIVAEHLUKIKPKSL----- 67
QY 70 VVEGPKRKRKPKRKAATQEKVKSKETGSAKKKNLKBESATKKPANVGDMSKNPEVTLKSCR 129
DB 68 -----RNPLQA-----KHKQNTSESSFAVLHSAESHKKQNYIPVKNKG 107
QY 130 ALNFDLENPGDARGDSESEIVQNSSGANSFSEIRDAIGTNGSFLDSVQIDKTNGLGA 189

Db 108 FTKQNGETPGIIAEEA-SKSE-----ESVSPKKPLFLQOPSELRRWRSEGA 151
QY 190 MNQPLEVS-MGNQPDKLGSTGAKLARDQOPDLLTRNQOC-----QFPVATQNTQF----- 237
Db 152 --DPAKFSDLEQDCSSLSKTRTD-----NSCISSHCGTTSPTSYSY-NTAFVLL 200
QY 238 -PMENQCAWLQMKNQLIGFFPGNQOPRMTIRN-----OQPCLAMGNQOPMYLIGTPRALV 292
Db 201 KAMEPELSTLSQKSPCAIKTEKLRPNKTKAPPKLANSMDAPNQISOELVAESQSCT 260
QY 293 SGNQQLGGPGNKRPIFLNHQTCPLPAGNQLYGSPTDMHQLVMSTGGQGHLLIKNQPGS 352
Db 261 SYTVHMSAAQKNE-----QGAQMSASHLY-----HQ-----HEHFVPKSNQH--- 297
QY 353 LIRGOQPCVPLIDQOPATPKGFT-----HLNQWATSM-SSPGL-RPHSOSQV 398
Db 298 -----NQOLPGCSGFTGSLTNLQENAKLEQVNIATVSVGLTSPSSRSQV 345
QY 399 PTVLHVESYRILNGTTGTCQSRAPAYDS-----LQODIHOGNKYILSHSISNG- 451
Db 346 TPQNOQMDASPLSISANSTQSPPMPIYNSTHVASVVSQVQMCNLLKLDQPKQKG 405
QY 452 -----CKKALPQNSLSLPTPIAKLEEARGSKRQVHRAMGQTEKHDLNLAQOIAQSDVER 506
Db 406 YICEYCNRAKAK-----PSVLLKHIRSHTGTPYPCVTCGFSFKTKSL-----YK 451
QY 507 HNSSTCVLEYLDAAKTKTIQVQENLHGM-----PPEVIEEDDPTD----- 548
Db 452 HKKS-----HAHTIKGLVLOPDAGGLFLESHPKALSTHSDVEDSGSEEGATDER 504
QY 549 ----GARKGNTASISKGAS-----KGNSSPVKKTAKEKCI 581
Db 505 QHDLGAMELQNVHIIKRNNAETLLKSSFTSPSPENVIGDFLLQDRSAEQAVTELKVV 564
QY 582 VPTPKAKKGAGRKKSVPVPAHAS-----EIQLWQTPPKTPPLSRKPKGKRKSTQDS 635
Db 565 VHRVTVSPLRTDSPKAMDPKPELSSAQKQDLQVTNVQPLSANMSQ-----GVSRLTNEN 621
QY 636 GKARGPSGELL-CQDS-IAETIYRMQNLVGLDKEREQE----- 671
Db 622 SHQKGMNPLEGGQDSHVGTVHAQLQEQATDYSQEQGKLLSPRSIGSTDSGVFSSSES 681
QY 672 -----QANVLYKGDGALVPYESKKRKRPR 696
Db 682 ADQTVSPPTPFARFPAQNKLEGTVDPLQLLSPRQHPLLCHREKALLP--GQMRPPLA 739
QY 697 KVIDIDTTIRIWNLLMGKDEKDEKDKKKKWEERR-----VFRGRADSFI 747
Db 740 TKTLERISK-----LISDNEALVDKQDLSVKPRRTSLSRGSDSPKSYIFK---DSF- 791
QY 748 ARMHLVQDRR-----FSPWKGVSVDVIGV-----FLTQNVSDHLSSSAFMSLA 792
Db 792 -QFDLKPVGRTSSSDIPKSPFTPTTEKSKQVLLSVPSLDCLPITRSNMTPTGYSAVP 850
QY 793 ARFPFKLSSSRREDERNVRSVVVEDPECCILNLEIPSQOEKV-QHPSDMEVSGVDSSKE 851
Db 851 ANIIPPPHPLRGSQ-----SFDOKIGAFYDDVVFSGPNA----- 884
QY 852 QLRDCNSNGIERNFLEKSIQNL-----EEVYLSQDSFDPALFQSCGRVSGSCSKSDAEF 908
Db 885 ---PVPOSGHPRTLVRQAAIEDSSANESHVLGTGQSLDES-HQGCHAAAG----- 929
QY 909 PTTRCETKTVSGTSQSVOTGSPNLSIDICLOGNERPHLYEGSGDVQKQETTNVAOKKPD 968
Db 930 -----EAMSVRSKALAQ-Phi-----EKKSHOGRGTWFECECTCRNRYK--L-970
QY 969 EKTMMNKDSCVFCQPRNDTNMQTTPSSSYEQCATRPHVLIDIEDFGMQGEGLYGS----- 1023
Db 971 ENFENHKKEVC-----SELGPKT-----KVAMREHSPVPG-GLQPIQLHYRVAGSS 1018
QY 1024 --WMSISPYDRVKNKNV---PRRFFQGGSVPREFTGQIIPTPHELPGMGLSGSSAV 1078


```
Db 1019 GIWEOTPIQIRKRRKMGVGGDDDELAQNESGTSPPKSSSEG-----LQFQNALGCNPSEL 1069
QY 1079 QEH-----QDDTQHOQDEMKASHLQKTFDLNLSSECLTQSSST--KONITDGCPLPRD 1132
Db 1070 PKHSVTIRSDQHKNIQLONSHIL-----VARGPEQTMDFPKLSTIMEQOISS----- 1117
QY 1133 RTAEDVVDPLSNSSLO-----NILVESNSSNKEQ-----TAVEKETNAT----- 1173
Db 1118 -AAQDKIELQRHGTGISVTHQNSLRSPNSFKPPEFRASPVSFQELNRTGNSGSLKVI 1176
QY 1174 -LIREMKGTADGKXP--TSOWDSLRL-----T----- 1196
Db 1177 GISQESHPSRDSHPHQALDALGELQESSRSPSRHVLGQPSRLIRHOHNQVPEI 1236
QY 1197 -----KDVEGNEGRQERNK-----NNMDSIDIYAI-----RRASISEISEAIKRG 1237
Db 1237 LVTEEPDRDLAQCHDOEKSEKFSQWSETLSKLPTEKLPKPKKRLRLAEIHSSTESS 1296
QY 1238 MNMLAVRI--KDFLERIVKDHGGIDLEWLRESPPDKADYL----- 1277
Db 1297 FOSTLSRSLRESSLSHTSSFSASLDIEDVSKTEASPKIDFLNKAFFLMIPAGLNTLNPV 1356
QY 1278 -----LSIRGLGLKSVEC-----VRLTLHLNLAFFVDNTNVCRIAV 1312
Db 1357 GCHREMRRTASEQINCTQTSMEVSDLSKSFDCGSGITPPQTPLTELQPP--SSPSRVGV 1414
QY 1313 RMGWVPL-----OPLPESLQ-----HLELYPVLESIQKFLWP----- 1346
Db 1415 -RGHPVLLERRRGPLVRQISLGIAPDLSHPVHPT--SFONTALPSVNAVPPYQGPOLTST 1471
QY 1347 -----RLC----- 1349
Db 1472 SLAEFSANTLHSGTQVQKLOAETSNSSTNVFPVQOLCDINLLNQIHAPPSSHQSTQLSLQ 1531
QY 1350 -----KLDQRTL-----YELHYQLITFGKVFCTKSRPNCGNACPMRGECRH 1389
Db 1532 VSTQSGKPKNSVLSSKSSSEDCFPAPKYLQHCQVFTSG-----PSCSSNPVH----- 1578
QY 1390 FASAYASARLALPAPEERSLTSATIP-----VPPESFP-----VPAIP-MIEL 1431
Db 1579 ---SLPNQVISDPVGTGDHCVTSATLPTKLIDSMNSHPLLPPLRPLGSGVQKVPSSFML 1635
QY 1432 PLPLEKSLASGAPSNRENCEPIIEFPASPGQCEITESEDIEDAYNEDDPIPIKL-- 1489
Db 1636 PIRLOASSVPAYCFATILSLPQLVLVQDLPNQIPICQTNHSWP---ISEQNSVPTLQKH 1692
QY 1490 -----NIEQFGM--TLREHMERNMELQEGDMSKALVALHPTTTSIPTPKLKNLSRLRTEH 1542
Db 1693 QNALPNPEKEFLCENVFSEMSONSLSE-----SLPITQISVGRLSPOQ 1737
QY 1543 QYVELPDSHRLDGMKREPDDPSPYLLAIWTPGETANSAPPEOKCGKSGKMC-FDE 1601
Db 1738 E--SSASSKRLM-----SPANSLDIAMEKHQKRAKDENGAVCATDV 1776
QY 1602 TCSEGNLSREANSQTVRGVLLI--PCRTAMRGSPFLNGTYFQWNELFADHESLKPIDV 1658
Db 1777 RPLEALSRVNEASKOKPILVRQVCTTE-----PLDGVMLE-KDVFQSPETSNEAVNL 1829

RESULT 9
DPOZ_HUMAN
ID DPOZ_HUMAN STANDARD; PRT; 3130 AA.
AC O60673; O43214;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DNA POLYMERASE ZETA CATALYTIC SUBUNIT (EC 2.7.7.7) (HREV3).
GN REV3L OR POLZ OR REV3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=98284025; PubMed=9618506;
RA Gibbs P.E.M., McGregor W.G., Maher V.M., Nisson P., Lawrence C.W.;
RT "A human homolog of the Saccharomyces cerevisiae REV3 gene, which
RT encodes the catalytic subunit of DNA polymerase zeta.";
RL Proc. Natl. Acad. Sci. U.S.A. 95:6876-6880(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow, and Leukocyte;
RX MEDLINE=99202263; PubMed=10102035;
RA Lin W., Wu X., Wang Z.;
RT "A full-length cDNA of hREV3 is predicted to encode DNA polymerase
RT zeta for damage-induced mutagenesis in humans.";
RL Mutat. Res. 433:89-98(1999).
RN [3]
RP SEQUENCE FROM N.A.
RC Murakumo Y., Rasio D., Roth T., Negrini M., Croce C.M., Fishel R.;
RT "Cloning and characterization of hREV3, the human homolog of S.
RT cerevisiae REV3.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 79-3130 FROM N.A.
RA Roth T., Rasio D., Murakumo Y., Negrini M., Croce C.M., Fishel R.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE OF 79-3130 FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=99126315; PubMed=9925914;
RA Morelli C., Mungall A.J., Negrini M., Barbanti-Brodano G., Croce C.M.;
RT "Alternative splicing, genomic structure, and fine chromosome
RT localization of REV3L.";
RL Cytogenet. Cell Genet. 83:18-20(1998).
CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -
CC N PYROPHOSPHATE + DNA(N).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -!- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED.
CC -!- DOMAIN: ITS C-TERMINAL PART COULD SERVE AS THE CATALYTIC DOMAIN
CC DURING NUCLEOTIDE POLYMERIZATION, WHILE ITS N-TERMINAL PART COULD
CC PROVIDE SITES FOR PROTEIN-PROTEIN INTERACTIONS WITH OTHER FACTORS
CC DURING TRANSLATION DNA SYNTHESIS.
CC -!- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-79 IS THE INITIATOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF058701; AAC24357.1; -
CC EMBL; AF071798; AAC24009.1; -
CC EMBL; AF157476; AAD40184.1; -
CC EMBL; AF035537; AAB88486.1; -
CC EMBL; AF078695; AAC28460.1; -
CC MIM; 602776; -
CC InterPro; IPR002064; -
CC Pfam; PF00136; DNA_POL_B; 2.
CC PRINTS; PR00106; DNA_POLB.
CC PROSITE; PS00116; DNA_POLYMERASE_B; 1.
KW Transferrase; DNA-directed DNA polymerase; DNA replication;
FT DNA-binding; DNA repair; Nuclear protein; Zinc-finger; Polymorphism.
FT ZN_FING 3042 3057 C4-TYPE (POTENTIAL).
FT ZN_FING 3086 3104 C4-TYPE (POTENTIAL).
FT VARIANT 231 231 Q -> H.
FT VARIANT 389 389 S -> T.
FT VARIANT 1540 1540 K -> E.
FT VARIANT 2607 2607 S -> T.
```

FT CONFLICT 237 237 /Frid=VAR_008519.
FT CONFLICT 1156 1156 E -> Q (IN REF. 4 AND 5).
SQ SEQUENCE 3130 AA; 352782 MW; 1C0700900F10BB14 CRC64;

Query Match 2.2%; Score 200.5; DB 1; Length 3130;
Best Local Similarity 18.3%; Pred. No. 0.02;
Matches 345; Conservative 229; Mismatches 647; Indels 663; Gaps 88;

QY 6 DSSAVNATEQNDGRQDVLFDLTKTQKPKRMPKVVVGGPKRKRPR-----61
Db 511 DNLASLSTIPQDGTADENSDNPLNENSRTHS-----SVIATSLSVKPSIFHK 562
QY 62 -----KPAELPKVVVEGPKRKRKAATQKVKSK-----TGSAKKKLKESAT-- 106
Db 563 DAATLEPSSAKITFOCK-----HTSALSHVLNKEDLIEDLSOTNKNTKGLDNSTVSF 617
QY 107 -----KRPANVGD-----SNKSPEVTLKSCRKALNFLEN--PGDARQGD-- 146
Db 618 TNESTYMKYPGSLSTVSHSENSHKENSKILLPVSSCESSI-FDYEDIPSVTRQVPSR 676
QY 147 -----ESELV-----QNSCANSESEIRDAIGTNGSFLDSVQIDKTNGLGA 189
Db 677 KYTNIRKIEKDSPIHMRHPNENTLGNFS-----NFDNLHNSKNKVS----- 720
QY 190 MNOPLVSMGNOPDKLSTGAKLARDQOPDLLTRNOQOCPVATQNTQPMENOQAWLQMK 249
Db 721 -----SEGKGNSTALSFL-----PSSEFTENCCELLSCSGENRTVMVHLSNSTADESL 769
QY 250 NOLIGFPFNOQPRMTIRNOQPCILAMGNOQPMYLLIGTPRALVSGNOOLGGP----- 302
Db 770 NKL-KIRVEEFOEHT--EKPSLSQAAHVMPF-----PSVVLN-CLTRPQKLSPTY 819
QY 303 -----GNKRPIFLNQTCLPAGNQLYGPTDMHQLVMGTGGQOGLLLKNQOPGLIRGQ 358
Db 820 KLOPGNK-PSRLKLNKRLAGHQBTSKSS-----BTGSKDNFIQNN----- 861
QY 359 PCVPLIDQOPATPKGFTHLQMVATSSPCLRPHSOSQVPTT-----YLHVESV 408
Db 862 PCNSNPEKDNALASDLTKTT-----RGAFENKTPDGFIDCHFGDGTLETEQ- 908
QY 409 SRLINGTTGQORSAPAY---DSLOQDIHOGNKYILSHISNGNG-----CKKALPON 459
Db 909 SFGLYGNKYTLRAKKVNYETEDSESSFVTHNSKISLPHPMEIGESLDGTLKSRKRKMS 968
QY 460 SSLPTPIIAK---LEEARGSKROYHRAMQOTKHD-----LNLAQOIAOSQDV--- 504
Db 969 KKLBP-PVLIKIYIINRRPRGRKNMLVK-LGKIDSKEKQVILTEEMKELYKKLAPLKDQFWPK 1026
QY 505 -----ERNSTTCVEYLDAAKT-KIQKVQENLHGMPPVEIEIEDDP 546
Db 1027 VPDSPATKPIYIPLTPKSHRRKS--KHSKAKKTKGQOQRTNENI----- 1070
QY 547 TDGARKGNKTASTKASKNGSPVKTT--AEKEKCIIVKTPAKKGAGRKKSVPVPPAHA 604
Db 1071 -----KRTLSFRKKRSHAILSPSPSYNAETEDCDLNTSDVMSKGLGLSERSTSPINS 1123
QY 605 SEIQWQOPTPKPTPLSRKPKGKRKSIQDSKARGSPGELLCCDSIAEIIYRMQNLV-- 662
Db 1124 SPPRCWSPTDPR-----AEEIMAAAEKAMLFKGPVNYKK 1158
QY 663 -----LGDKEREQONAMWLYKGDGALV--PYBSKKRPRPKVIDDETTRINLLMGK 715
Db 1159 TVNSRIGKTSRRAQ-----IKSKAKLANPSIVTKKNNR-----NOTKNLVD----- 1202
QY 716 DEKEGDEKKKEKWEERVRFRGRADSFIAHMLHVQDGRFRSPKWSVDSVIGVFL 775
Db 1203 DGKKKPRAKOKTNEKG-----TSRKHITLADKIKSOGSAEVKFLV---K 1244
QY 776 TONVSDHLSSAFMSLAAREFPKLSRSSREDERNVRSVVVEDPECCILNLNEIPSWQEKVQ 835
Db 1245 HONVSEFASGGSGSL-----LFFKQKDMPLMGSAVD 1275

QY 836 HPSDMEV-SGVDSGSKBQLRDCNSGIERFNFLEKSIONLEEEVLSSQDSFDPALFOSCG 894
Db 1276 HPLSASLPTGIN-AOOKLSGCFSSFLS-----KKSVD--LQTFPSSRDDLHRSV--CN 1325
QY 895 RVGSCSKSDAEFPPTTRC-----ETKTVSGTSSQSVQT 927
Db 1326 SIGP-GVSKINQVRPHOSAMFTLKSTLIQKNIFDLNHLSQLVAQNTQISSGSSKIED 1384
QY 928 GSNLSDICLQGNRPHLYEGSDVQKQETTNNVAOKKPDLEKTMW---KDS-----VCF 980
Db 1385 NANNI-----QRNYL---SSIGKSEYNSLESKLDQAYTPNLFHCKDSQQQIVCI 1432
QY 981 GQ-----PRNDTN-----WOTTPSSSYECATRQPH---VLDIEDFGMGE 1019
Db 1433 AEOGKHSETCPGNTASEESOMPNNCFVLSRPIKIAWEQKQKRGFILDMSNFKPE--- 1489
QY 1020 LGYSWMSISPR-----VDRVKNVPRFRFGGSGVPREFTGQIIPSTPHELPG 1068
Db 1490 -----RVKPRSLSEALSQTKALSQCKNRV-----STPSAF-----G 1521
QY 1069 MGLSGSSSAVQEHQDDTOHNOQDEMKNASHLQKTFLLDLNLSSECLTRQSTKQNTIDGC 1128
Db 1522 EGOSG-----LAVKELLOKROOKAQN----- 1543
QY 1129 LPRDRTAEDVVDPLSN---NSSILQNLVESNSNKEOTAVEYKETNATILREMKGTAD 1184
Db 1544 -----ANTQDPLSNKHOPKNKISGL-EHNKANKTRSV-----TS 1579
QY 1185 GKPTOSMDSLRKDVEGNEGRQERNKNMDSIDYAIRRASISEISAIRKERNMMLAV 1244
Db 1580 PRKPRTPRSTKQKE-----KIPKLLKVDLSN---L 1606
QY 1245 RIKDFLERIVKDHGGIDLEWLRSPPDKAKDYLLSIRGLKSGVECVRLTLHLNAPVD 1304
Db 1607 QNSQLDNSVSDSPISFF-----SDPGFESCYSL-----DLSLSPHNYNFDIN 1650
QY 1305 T--NVGRIARVMG--WVPL-QPLPES-LQLHLELYPVLESIOK--FL----- 1344
Db 1651 TIGOTGFCFSYSGSQFVPADQNLQKFLSDAVQDLFGP-QAIEKNEFLSHDNQKCEDKH 1709
QY 1345 -----WPLKCLD-----QRTLYELHYQLITFGKVFCTKSRPN 1377
Db 1710 HTTDSASWIRSGTSLPSEIFEKSTIDSNENRRHQKNKNSFPLTRSNISIMDSFCVQOAE 1769
QY 1378 CNACPMRGECHRFASAYASARLALPAPEE-----RSLTSAT-----IPVPP 1418
Db 1770 CLSEKSR---LNRSSVSKEVFLSLPQPNNSDWIQGHTRKEMGQSLDSANTSFATLSPD 1826
QY 1419 ESPPPVAIPMIEL-----PLPLEKSLAGAPSNNR-----CEPIIEPA 1458
Db 1827 GELVDVACEDLELVSRNNDMLTPTDSSPRSTSSPSQSKNGSFPTPTANILKPLM-SPP 1885
QY 1459 SPOECTEITESDIEDAYNE---DPDEIPTKLNEQFGMTLREHMERNMELQGDMS 1514
Db 1886 SREIMATLDDHLDSETIYOEPFCNSPNPVEKPREITGGRLMAYETRLANDLAEFEGDFS 1945
QY 1515 KALVALHPTTSTPTPKLKNISRLRTEHOYVELPDSHRLLDGMDKRE---PDDPSPYL 1569
Db 1946 LEGRLWKTAFASTQNPGRPSLRSGQVYVNGSSNPKMWEDKKIVIMPCKAPSRQL 2005
QY 1570 LAIWTEG--ETANSAPPEQKCG 1591
Db 2006 VOYWLQAKEEYERSKLPKTKPTG 2029

RESULT 10
AKAC_HUMAN
ID AKAC_HUMAN STANDARD; PRT; 1781 AA.
AC Q02952; Q99970; Q00498; Q00310;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)


```
QY 730 KWEBEERRVFRGRADSFIAHMLVQGDRRFSPWKGSVVDSVIGVFLTONVSDHLSAFA 789
Db 639 -----SALISSTE-- 646
QY 790 SLAARFPKLSRSSREDNRVSVVDEPECIINLNLIPISQEKV-----QHPS 838
Db 647 STASEQEMKGSVEEPK-----PEPKRKVDTSVSWEALICVSGSKRRARRSS 697
QY 839 DMEVSGVDSGSKQLRDCNSGIERFNFLKSTONLEEEVLSSQSDSDPAIFQSCGRVGS 898
Db 698 DEEGGPKAMGDGHOKAD--EAG-----KDKETGTDGLAGSQEHPD-----GQ 738
QY 899 CSKSKSDAEFPPTTCETKTGTSQSQTGSPNLSDBICLQGNRPHLYEGSGDVQKQET 958
Db 739 GSSSPEQAGSPT---EGEGVSTWESFKRLVTPRKSKSKLEEKSEDSI--AGSG----- 787
QY 959 TNVAQKPKDLKTNWKNKDSVCFQOPRNDTNW-----OTTPSSSYEQCATRPHV 1007
Db 788 --VEHSTPDTE-----PGKEESWVSIKFPGRKKRPDPGKQEQ----- 825
QY 1008 LDIEDFCMQGGLGYSWM--SISP--RVDRVKNKNVPRFFRQGSVPR-----EFTGQII 1059
Db 826 -PVEDAGTGANEDSDVPAVVPVSEYDAVEREKMEQAQKGAQPEQKAAVEVSKELS 884
QY 1060 PSTPHELPGMGLSGSSA--VQEHQDDTOHNOQDMNKASHLQKTFDLNLSSEECITRQ 1117
Db 885 ESQVHMAAAVADGTRAATIIERSPSN-----ISASVTEPLEQV 924
QY 1118 SSTKQNTDGLCPDRPAED-----VVDPLNSSLQ--NILVSNSSNKKOTAVEYKTN 1171
Db 925 EAAEALLTEEVLEEVTAEEBPTVTPELPENREARGDTVVSAAELTPEAVTAAE---T 980
QY 1172 ATILREKMGTLADGKKPTSDSL-----RKDVEGNEGQERKNKNMDS 1215
Db 981 APLGSEGETASAAETEMWSAVSQLTDSPTTEATPVQVEGVPDIEQERTQE 1040
QY 1216 IDYBAIRRASISEISEAIKER---GMNNMLAVRIKDFLERIVKDHGGIDLEWLRESPPD 1271
Db 1041 V-----LQAAVKEESQSLPGTGGP-----EDVLQPVQRAEA-----ERPEE 1078
QY 1272 KADYLLSIRGLGKSVCEVRLTLNLAFVDTNVGRI--AVRMGWVPLOPLPESLQHL 1330
Db 1079 QAE-----ASGLKET-----DVVLKVDQAQEAKEPFTQGVVGGTTPSEF--- 1120
QY 1331 LELPVLESIQKFLWPLRLKLDORTLYELHYLITFGKVFCTKSRPNACPMRGEGRHF 1390
Db 1121 -KAPQVTESES-----SELV-----TCQAEFLAG----- 1145
QY 1391 ASAYASARLALPAPEERSLTSATIPVPPFPVPAIPMIELPLP-----LEKS 1438
Db 1146 ---VKSQEMVM---EQAIIPDSVEITDSETDGTSTPVADFGPTTKDVEIVEIHEENE 1198
QY 1439 LASGAPSNRENCEPI---IEEPASG---OECTETESDIEDAYVNEPDP-EIPTKL 1489
Db 1199 VASGTQSGGTEAAVPAQKRPAPSPFVQEEK-BQSKMEDTLEHTDREVSVEIVSI 1256
RESULT 11
BRCL_MOUSE STANDARD; PRT; 1812 AA.
AC P48754; Q60957; Q60983;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE BREAST CANCER TYPE 1 SUSCEPTIBILITY PROTEIN HOMOLOG.
GN BRCL1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=C57BL/6; TISSUE=Embryo;
RX MEDLINE=96177659; PubMed=8634697;
RA Abel K.J., Xy J., Yin G.Y., Lyons R.H., Meisler M.H., Weber B.L.;
RT "Mouse Brca1: localization sequence analysis and identification of
RL evolutionarily conserved domains.";
RN Hum. Mol. Genet. 4:2265-2273(1995).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RX MEDLINE=96177660; PubMed=8634698;
RA Sharan S.K., Wims M., Bradley A.;
RT "Murine Brca1: sequence and significance for human missense
RL mutations.";
RN Hum. Mol. Genet. 4:2275-2278(1995).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RX MEDLINE=96121367; PubMed=8575748;
RA Bennett L.M., Haugen-Strano A., Cochran C., Brownlee H.A.,
RA Fiedorek F.T. Jr., Wiseman R.W.;
RT "Isolation of the mouse homologue of BRCA1 and genetic mapping to
RL mouse chromosome 11.";
RN Genomics 29:576-581(1995).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ; TISSUE=Embryo;
RX MEDLINE=96067162; PubMed=7590247;
RA Lane T.F., Deng C., Elson A., Lyu M.S., Kozak C.A., Leder P.;
RT "Expression of Brca1 is associated with terminal differentiation of
RL ectodermally and mesodermally derived tissues in mice.";
RN Genes, Dev. 9:2712-2722(1995).
[5]
RP SEQUENCE OF 727-1111 FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Embryo;
RX MEDLINE=96021028; PubMed=7550308;
RA Marquis S.T., Rajan J.V., Wynshaw-Boris A., Xu J., Yin G.Y.,
RA Abel K.J., Weber B.L., Chodosh L.A.;
RT "The developmental pattern of brca1 expression implies a role in
RL differentiation of the breast and other tissues.";
RN Nat. Genet. 11:17-26(1995).
[6]
RP SEQUENCE OF 789-1250 FROM N.A.
RC STRAIN=129/SVJ;
RX MEDLINE=96163506; PubMed=8566965;
RA Schroeck E., Badger P., Larson D., Erdos M., Wynshaw-Boris A.,
RA Ried T., Brody L.;
RT "The murine homolog of the human breast and ovarian cancer
RL susceptibility gene brca1 maps to mouse chromosome 11D.";
RN Hum. Genet. 97:256-259(1996).
CC -!- FUNCTION: NOT KNOWN, MAY REGULATE GENE EXPRESSION. INVOLVED IN
CC TRANSCRIPTIONAL REGULATION OF P21 IN RESPONSE TO DNA DAMAGE (BY
CC SIMILARITY).
CC -!- SUBUNIT: CIP1 INTERACTS SPECIFICALLY WITH THE BRCT DOMAINS (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -!- TISSUE SPECIFICITY: IN THE EMBRYO, EXPRESSED IN OTIC VESICLES AT
CC DAY 9.5. AT DAY 10.5, THIS EXPRESSION DECREASES AND HIGH LEVELS
CC ARE FOUND IN THE NEUROECTODERM. AT DAYS 11-12.5, HIGH LEVELS IN
CC DIFFERENTIATING KERATINOCYTES AND WHISKER PAD PRIMORDIA. AT DAYS
CC 14-17, EXPRESSION ALSO OBSERVED IN KIDNEY EPITHELIAL CELLS. IN
CC THE ADULT, HIGHEST LEVELS FOUND IN SPLEEN, THYMUS, LYMPH NODES,
CC EPITHELIAL ORGANS, AND ALVEOLAR AND DUCTAL EPITHELIAL CELLS OF
CC THE MAMMARY GLAND. VERY LOW LEVELS IN BRAIN, KIDNEY, AND SKIN. NO
CC EXPRESSION IN HEART, LIVER OR LUNG.
CC -!- DEVELOPMENTAL STAGE: IN THE MAMMARY GLAND, EXPRESSION INCREASES
CC DRAMATICALLY DURING PREGNANCY. LEVELS FALL DURING LACTATION AND
CC INCREASE AGAIN DURING POST-LACTATIONAL REGRESSION OF THE
CC MAMMARY GLAND.
CC -!- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
CC -!- SIMILARITY: CONTAINS 2 BRCT DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```


Qy 1216 IDYEAIRRASISE-----ISAIKRGMMNNLAVRIKDFLERIVKDHGGIDLEWLRSP 1270
Db 1601 TPKEVMTOKFAERYKRLTLTAITEETHVIKTKDAEVCERTLK-----1645
Qy 1271 DKADYLLSING-----IGLKSVECVRLTLTHNLAFFVDVTNVR 1309
Db 1646 -----YFLGIAGGKWIWSYVWVRSIOERRLLNHEFEVTGDVVTGR 1687

RESULT 12
CBP_HUMAN
ID CBP_HUMAN STANDARD; PRT: 2442 AA.
AC Q92793; Q16376; O00147;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE CREB-BINDING PROTEIN.
GN CREBBP OR CBP.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=97385172; PubMed=9238046;
RA Sobulo O.M., Borrow J., Tomek R., Reshimi S., Harden A.,
RA Schlegelberger B., Housman D., Doggett N.A., Rowley J.D.,
RA Zelenik-Le N.J.;
RT "MLL is fused to CBP, a histone acetyltransferase, in therapy-related
RT acute myeloid leukemia with a t(11;16)(q23;p13.3).";
RL Proc. Natl. Acad. Sci. U.S.A. 94:8732-8737(1997).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=97321049; PubMed=9177780;
RA Giles R.H., Petrij F., Dauwerse H.G., den Hollander A.I.,
RA Lushnikova T., van Ommen G.J.B., Goodman R.H., Deaven L.L.,
RA Doggett N.A., Peters D.J.M., Breuning M.H.;
RT "Construction of a 1.2-Mb contig surrounding, and molecular analysis
RT of, the human CREB-binding protein (CBP/CREBBP) gene on chromosome
RT 16p13.3.";
RL Genomics 42:96-144(1997).
[3]
RN SEQUENCE OF 1-405 FROM N.A.
RX MEDLINE=96376968; PubMed=8782817;
RA Borrow J., Stanton V.P., Andresen J.M., Becher R., Behm F.G.,
RA Chaganti R.S.K., Clavin C.I., Distche C., Dube I., Frischauf A.M.,
RA Housman D., Miteiman F., Volinia S., Watmore A.E., Housman D.E.;
RT "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses
RT a putative acetyltransferase to the CREB-binding protein.";
RL Nat. Genet. 14:33-41(1996).
CC -1- FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO
CC PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP AUGMENTS
CC THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF
CC CAMP-RESPONSIVE GENES.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DISEASE: INVOLVED IN ACUTE LEUKEMIAS BY CHROMOSOMAL TRANSLOCATIONS
CC T(8;16)(p11;p13) INVOLVING CBP AND MOZ, AND T(11;16)(Q23;P13.3)
CC INVOLVING CBP AND MLL.
CC -1- DISEASE: DEFECTS IN CBP ARE THE CAUSE OF RUBINSTEIN-TAYBI
CC SYNDROME (RSTS), A DISORDER CHARACTERIZED BY CRANIOFACIAL
CC ABNORMALITIES, BROAD THUMBS, BROAD BIG TOES, MENTAL RETARDATION
CC AND A PROPENSITY FOR DEVELOPMENT OF MALIGNANCIES.
CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; U47741; AAC51770.1; -;
DR EMBL; U85962; AAC51331.1; -;
DR EMBL; U89354; AAC51339.1; -;
DR EMBL; U89355; AAC51340.1; -;
DR MIM; 600140; -;
DR InterPro; IPR000197; -;
DR InterPro; IPR000433; -;
DR InterPro; IPR001487; -;
DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF02135; zf-TAZ; 2.
DR PRINTS; PR00503; BROMODOMAIN.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS00633; BROMODOMAIN_2; 1.
KW Transcription regulation; Nuclear protein; Activator; Bromodomain;
KW Chromosomal translocation.
FT DOMAIN 363 430 CYS/HIS-RICH.
FT DOMAIN 452 683 CREB-BINDING.
FT DOMAIN 1103 1175 BROMODOMAIN.
FT DOMAIN 1061 1064 POLY-GLU.
FT DOMAIN 1199 1487 CYS/HIS-RICH.
FT DOMAIN 1555 1562 POLY-GLU.
FT DOMAIN 1675 1849 CYS/HIS-RICH.
FT DOMAIN 1943 1948 POLY-PRO.
FT DOMAIN 2081 2085 POLY-GLN.
FT DOMAIN 2199 2216 POLY-GLN.
FT DOMAIN 2245 2248 POLY-GLN.
FT DOMAIN 2297 2300 POLY-GLN.
FT CONFLICT 1511 1513 FAE -> NSG (IN REF. 2).
FT CONFLICT 1724 1725 ED -> VV (IN REF. 2).
FT CONFLICT 1770 1770 V -> L (IN REF. 2).
FT CONFLICT 1789 1789 N -> F (IN REF. 2).
FT CONFLICT 1812 1812 T -> P (IN REF. 2).
SQ SEQUENCE 2442 AA; 265336 MW; 42D084619475F3D2 CRC64;

Query Match 2.1%; Score 192; DB 1; Length 2442;
Best Local Similarity 18.1%; Pred. No. 0.04;
Matches 345; Conservative 238; Mismatches 720; Indels 598; Gaps 89;

Qy 5 MDSSAVNATEATONGSDRODVFEDLNKT-----POKPSKRRKRPMPKVVGEKPKRK 59
Db 288 MGATGVNPQLASKQSMVNSLPTFTDIKNTSVTPNNSQMOTSVGIVPTQAIATGPTAD 347
Qy 60 P--RKPALPKVVVEGPKRPKRAATQEVKSKETGSAKKKLNKESATK----KPAVNG 113
Db 348 PEKRLTQQQLLVLLHAKCORREQANGEVRACSLPHCRMTMKNVNLNHTHCQAGKACQVA 407
Qy 114 DMSNKSPEVT-LKSCRK---ALNFDLENPGDARQGD-----SESEIVQSSGSAFSFIRD 165
Db 408 HCASSROIISHWKNCTRHDCPVCLPLKNASDKRNQQTILGSPASGIQNTIGS-----V 460
Qy 166 AIGTNGSFSDVSQIDKTN-----GLGAMNQPLEV-----196
Db 461 GTGQGNATLSNPIDPSSNQRAYAALGLPYMNPOTQLQPQVPGQPAQPOTHQOMRT 520
Qy 197 --SNGNOPDLSTCAKLARDQDPLTRNQOCOPPVATQNTQPMENQOAWLQMKNLIG 254
Db 521 LNPGLNPMNIPAGG-ITTDQPPNLLISEALPTSIGATN---PLMND-----G 565
Qy 255 PFPGNQPRMTIRNQQLCLAMGNOQPMY-----LIGTPRPA-----290
Db 566 SNSNGIGTSLTIPTAAPSPSTGVGRKGWHEVTDLRSHLVHKLVAQIFPTDPAAKDDR 625
Qy 291 ---LVSGNQQLGG---PQGNKRPIFLNHQCLCPAGNOLYGSPTDMHOLVMSTGGQHGILL 344
Db 626 MENLVAYAKVEGDMYSEANSRDEYH-----LLAEKIYKIQKELEEKRS---RLKQG 677
Qy 345 IKNOQPSGLRGOQPCVPLDQ-OPATPKGFTHLNQWVATSMSSPGLRPHRSQOVPTYL 403
Db 678 ILGNQPALPAPGAQP--PVTPAQQVRP-----PNGSLSLPVNRM 715

QY 404 HVESYRILNG-TTGTGCRSAPAYDSLOODIHQGNKYLISHSISNGNGCKKALPQNSSL 462
 Db 716 QVSOQMNFSNLSGNVQLPOAPM-----GPRASPMNHVS 751
 QY 463 PTPIMAKLEEARGSRQYHRAMQTEKHDNLN-LAQOIAQSQDVERHNSSTCWEYLDA--- 518
 Db 752 QMNSGVPGMAISPRMPQPPNMGAHTNNMMAQAPASQFLPQNFPSGSGAMSVGMG 811
 QY 519 -----AKTKIKQVVOENLH----- 533
 Db 812 OPPAQTVSQGVPGAALPNLMLGPOASQLPCPPVTSQPLHPTPPASTAAGMPSLQH 871
 QY 534 ---GMPPEVIEIDPTDGARKGNTASISKG-----ASKGNSSP-VKKTAEKEKCIYVPK 584
 Db 872 TTPPGMTTPQPAAPTQPTSPVSSSGQTPTPTPGSVPSATQSTPTVQAAAQAQVTPQPO 931
 QY 585 TPAGKGRAGRKSS---VPPPAH-----ASEIQLWQTP-----PK 616
 Db 932 TPVQPPSVATPOSSQOQTPVHAQPPGTPLSQAASIDNRVPTPSSVASAETNSOQPGPD 991
 QY 617 TPLSRKPKGKRKSIQDSGKARG-PSGELLQDSIAEIIYRMQNLXILGDKEREQEONAM 675
 Db 992 VPVLEKMTQAEATEDPDGSGESKGPRESMEMED-----LQASQVKEETDI 1038
 QY 676 VLYKGDGALVPYSEKRPKPRKVIDDDETRINWLLMGKDEGEDEKDKKKKWEWEE 735
 Db 1039 ABQKSE-----PMEVDEKPEVKEVEKEEESNN---GTASQSTSPSPQPKKIFK-PEEL 1090
 QY 736 RVFRGRADSFATRHLYVQGRDRFRFPWKGSVVDVSVIGVFLTONVSDHLSSSAFMSLAARF 795
 Db 1091 RQALMPTLEA-----LYRQDPESLPFPQPDQLLGI---PDYEDIVKNPMDLSTIKR- 1140
 QY 796 PPKLSSREDE---RNRSVVVDPPGCIILN--LNEIPSGQKRVQHPDSMEVS-----GV 845
 Db 1141 --KLDGTQYQEPWQYVDDVLMFNNAWLYNRKTSVYKFCSLAEVFEQIEDVPMQSLGY 1198
 QY 846 DSGSK-----EQL-----RDCNSGIE-RFNLEKSIQNLSEVLS----- 880
 Db 1199 CGGRKYEFPQTLCCYQKQOLCTIPDAAYSYONRYHFCEKCFTEIQGENVTLGGDDPSQ 1258
 QY 881 -----SODSFPAIFQSGCRVGS-----C-SCRSKDAEF 908
 Db 1259 QTTISKQDFEKKKNTLDPEPPVDCCKGRKMHQICVLHYDIWPSGFCVDCNCLKTKGR- 1317
 QY 909 PTTRETCTVSTGTSQVQSGPNLSEI--CLOGNERPHLYEGSDV-----QKQETTN 960
 Db 1318 --PRKENFSARKLQTLRLGN-HLEDVYNKFLRRQNHPE-----AGEVFVYVASSDKTVE 1370
 QY 961 VAOKKPDLEKTMNWKDVCFCQPRNDTNQWITPSSSYEQCATRQPHVLDIEDFGWQEGCL 1020
 Db 1371 V---KPGMKS---RFVDS---GEMSESFYRTKALFAFEEI-----DGVDVCFGMHVQY 1417
 QY 1021 GYSWMSISPRVDRVKNKNVPR-RFRFGQSVPR-----EF-----TGO 1057
 Db 1418 GSD--CPPNTRRVISYLDSTHFR-----PRLRTAVYHILIGLEYVKKLGVYVGH 1470
 QY 1058 ITPSTPHLPGLMGSGSSAVQEHQDDTQHNQODDMKASHLQKTFDLNLNS--SEECUT 1115
 Db 1471 IWACPPSE-----GDDYIFHCHPDD-----QKIPKPKRLQENYKMKLDKAFERI 1517
 QY 1116 ROSSFTQNTDCLPRDRTAEDVDVPLNNSSLQNLIVESNNSKEQTAVEYKFNATIL 1175
 Db 1518 DYKIDFKQATE-----DRUTSAKELPYFEGDPWPVNLVEESIKELEQEEERKEESTAAS 1572
 QY 1176 REMKGTADGKKPTQOWDSLRKDVEGNEGRERNKNMDSIDYEAIRASISEISEATKE 1235
 Db 1573 ETTEGSGDSK-----NAKKKNKTKNNK-----SSIRANKK 1606
 QY 1236 R---GMNNLAVRIKDFLER-----IVKHGGIDLEWLRESPPDKADYLLISIRGLGL 1285
 Db 1607 KFSMPNVSNLQSLKYATMEKHEVFFVHLHAGPVINTL---PPIVDDPDLSS----- 1657
 QY 1286 KSVCEVRLLTLNLNAPPVDTNVGR-----IAVRMGWVLPQLPESLQLHLLELYPVLES 1340

RESULT 13

TPR_HUMAN
 ID TPR_HUMAN STANDARD; PRT; 2349 AA.
 AC P12270;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE NUCLEOPROTEIN TPR.
 GN TPR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93064711; PubMed=1437155;
 RA Mitchell P.J., Cooper C.S.;
 RT "The human tpr gene encodes a protein of 2094 amino acids that has
 RT extensive coiled-coil regions and an acidic C-terminal domain.";
 RL Oncogene 7:2329-2333(1992).
 RN [2]
 RP REVISIONS, AND CHARACTERIZATION.
 RX MEDLINE=95096166; PubMed=7798308;
 RA Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,
 RA Saphire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;
 RT "tpr, a large coiled coil protein whose amino terminus is involved in
 RT activation of oncogenic kinases, is localized to the cytoplasmic
 RT surface of the nuclear pore complex.";
 RL J. Cell Biol. 127:1515-1526(1994).
 RN [3]
 RP SEQUENCE OF 1-142 FROM N.A.
 RX MEDLINE=88262257; PubMed=3387099;
 RA King H.W.S., Tempest P.R., Merrifield K.R., Rance A.J.;
 RT "tpr homologues activate met and raf.";
 RL Oncogene 2:617-619(1988).
 CC -!- FUNCTION: COMPONENT OF THE CYTOPLASMIC FIBRILS OF THE NUCLEAR PORE
 CC COMPLEX IMPLICATED IN NUCLEAR PROTEIN IMPORT. ITS AMINO TERMINUS
 CC IS INVOLVED IN ACTIVATION OF ONCOGENIC KINASES.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF THE NUCLEAR PORE
 CC COMPLEX. THE ASSEMBLY OF THE NPC IS A STEPWISE PROCESS IN WHICH
 CC TRP-CONTAINING PERIPHERAL STRUCTURES ASSEMBLE AFTER OTHER
 CC COMPONENTS, INCLUDING P62.
 CC -!- TISSUE SPECIFICITY: HIGHEST IN TESTIS, LUNG, THYMUS, SPLEEN AND
 CC BRAIN, LOWER LEVELS IN HEART, LIVER, AND KIDNEY.
 CC -!- DISEASE: INVOLVED IN TUMORIGENIC REARRANGEMENTS WITH THE MET, TRK
 CC OR RAF GENES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

QY 1497 TLREHM-----ERNMELQEGMDKALVALHPTTTSITP-
DB 1819 VSATPSSSLPKRTREEDSTIEASQVSDTVMPLPKLASVTPVGTEEV

RESULT 14

ATI_COMPX STANDARD; PRT: 1284 AA.
AC PI6602;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1990 (Rel. 16, Last annotation update)
DE A-TYPE INCLUSION PROTEIN (ATI).
GN ATI.
OS Cowpox virus (CPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10243;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CPRO6;
RX MEDLINE=88089536; PubMed=2826668;
RA Funahashi S., Sato T., Shida H.;
RT "Cloning and characterization of the gene encoding the major protein
of the A-type inclusion body of cowpox virus.";
RL J. Gen. Virol. 69:35-47(1988).
RN [2]
RP SEQUENCE OF 1-109 FROM N.A.
RX MEDLINE=88111568; PubMed=2828037;
RA Patel D.D., Pickup D.J.;
RT "Messenger RNAs of a strongly-expressed late gene of cowpox virus
contain 5'-terminal poly(A) sequences.";
RL EMBO J. 6:3787-3794(1987).
CC -!- FUNCTION: MATURE VIRUSES ARE OCCCLUDED INTO THE ATI, AND IT HAS
BEEN ASSUMED THAT SUCH BODIES PROTECT THE VIRUS DURING
DISSEMINATION FROM ANIMAL TO ANIMAL.
CC MISCELLANEOUS: A CHARACTERISTIC FEATURE OF ATI IS THE FORMATION OF
LARGE MASSES WITH NO SURROUNDING MEMBRANES IN THE CYTOPLASM OF
INFECTED CELLS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D00319; BAA00222.1; .
DR EMBL; X06343; CAA29650.1; .
DR PIR; JQ0006; WWVZAI.
KW Late protein; Repeat.
FT DOMAIN 611 912 10 X APPROXIMATE TANDEM REPEATS.
FT REPEAT 611 637 1.
FT REPEAT 638 665 2.
FT REPEAT 666 689 3.
FT REPEAT 690 720 4.
FT REPEAT 721 751 5.
FT REPEAT 752 780 6.
FT REPEAT 781 811 7.
FT REPEAT 812 842 8.
FT REPEAT 843 871 9.
FT REPEAT 872 912 10.
SQ SEQUENCE 1284 AA; 150329 MW; F7904C9E1DE8D012 CRC64;

Query Match 2.1%; Score 190; DB 1; Length 1284;
Best Local Similarity 18.4%; Pred. No. 0.021;
Matches 223; Conservative 159; Mismatches 387; Indels 442; Gaps 56;
QY 469 KLEEARGSROYHRANGOTEKH-----DLNLAQOIQSQDVERHNSSTCVYLDAAKTKI 524

DB 414 KCEEECGNGEYKTELEAKRKLTLELNLSDKLSKIRTLE-----DSVYKT-- 461
QY 525 QKVQENLHGMPEVIEIEDPTDGARKGKNTASISKGASKGNSSPVKTAKEKCIIVPK 584
DB 462 -----ERIDRLTKEIKELRD-----IQNTDGGSDSSEIDKKTIRELRESLDR----- 506
QY 585 TPAKGRAGRKKSVPVPAHASEIQLWQTPPKTPLSRKPKGKGRKSIQDSKGARPSGE 644
DB 507 -----EMTELERE-----LDIRD-GKVEGS----- 527
QY 645 LLCQDSIAEIIYRMQNLVYLGDKEROENAMVLYKGDGALVPYESKKRPRKPRKVDDEET 704
DB 528 --CQ-----RELESRMWL-----KORDDDLRAEIDKRR 554
QY 705 TRIWNLGMKGDEKED---EEKDKKE-----KWEEERRVFRGRASDF-IARMH 751
DB 555 NVEWELSLRLDIKEDKYKEDLDKAKTTISNVVSRISTLESIKAYQODRDTLSVVRRE 614
QY 752 LVQGDRRFSPWKSVDVSVIGVFLTONVSDHLSSSAFMSLAARFPKPLSSSREDERNVRS 811
DB 615 LEEERRVRDLERLDE-----CTRQEDTOEVDALRSRIRELENKLRADCMESGGG--- 665
QY 812 VVVEDPEGCILNLEIPSWQEKVQHPDMEVSGVDSGSKQEQLRDCNSGIERENFLEKSI 871
DB 666 -----NLTEISRLQSKI---SDLE-----RLRECRGNATE-ISRLQYRI 701
QY 872 QNLEEVLSQDSQDPAIFQSGRGVSGSCSKSDAFPTTRCTETKTVSGTOSVQTGSPN 931
DB 702 TDLERQL-----NDCRRNE---NNADTEREMQRLDRITDLERQ----- 738
QY 932 LSDEICLOGNERPHLYEGSDYQKQETTNNVQAKKPPLEKTMNWKDSVCQGPNDTNWQT 991
DB 739 LSD--CRRNN-----ESNADMER-EMQRLDRIMDLDRQLN----- 771
QY 992 TPSSSYEOCATQPHVLDIEDFGMGEGLGYSWMSTSPRVDRVKNVPRRFRPGGSGVP 1051
DB 772 -----EC-----KDGNGTS-----SEVNRLKTRI----- 792
QY 1052 REFTGOIIPSTPHELPGMLSGSSSAVQHQDDTQHNOQDEMKNASHLOKTFDLNLSSE 1111
DB 793 -----RLERSLEICSDSELYSAYKSELGRAR-----EQISNLQ 828
QY 1112 ECLTROSSTKQNTDCCLPRDRTAEDVDPVPLSNSSQLNLYVESNSS----- 1158
DB 829 ESLRERESDK--TDSYYRELTRE-----RNKIVELEKELAKCFDTNHAKYI 874
QY 1159 ---NKBOTAVEYKETNATILREMKGTADGKKPTQOWDSLKRDV---EGNEG----- 1204
DB 875 DEINSKTRISDLEROLAAACKNGSGNDMDQVKRIESLKLRELAECRGNNGSHSDCKY 934
QY 1205 -----RQE---RNKNMNDSDIYEAIRRASISEISEAIKERGMNMLA--VRKIDFLERIVK 1255
DB 935 YDEEAREEVKRLRQELTQL-HEDLKRARESDKNDYSYKRELEQRQAKVIEVEKELERYD 993
QY 1256 D-----HGGIDLEWLRESPPDKAKYLLSIR--GLGLKSVCEVRLTLHLNLAFPVD 1305
DB 994 DSRLAECRKHGD---EMLR-----KIADLEKLRDGGNGGNGCTSSCEFER----- 1038
QY 1306 NVGRIAVRMGWVPLQPLPESLOHLLELYPVLESIQKFL-WPRL---C--KLDO----- 1353
DB 1039 ---KRIAV-----LEVEVKSMETIKSLKFMFEDRLQKDCADKLDREKERM 1083
QY 1354 RTLYELHYQITFGKVCTKSRPNCNACPMRGECRHFASAYASARIALPAPEE----- 1406
DB 1084 KAERDLEREI-----ARKNCGGNPCERELESERSNVKRLEYQVDAEKVKFYKR 1133
QY 1407 -----RSLSATIPVPPSPPPVPAIPMIELPLPLEKSLASGAPSNRCEPIIEE 1456
DB 1134 ELERDYLSSRYLTSSSDP-DEKPLPNYTFPRIEV-----EPLTTDETPEKPYE 1181
QY 1457 PASPOECTEITESDIEDAYNEDPEITIKINTIQFGMTLREHMERNMELQEGDMSKA 1516
DB 1182 VVPPSSDVTEPISSGV-----TPSDAEPE-----HPOLSEYQTSVS 1218

QY 1517 LVALHPHTTSTPKLKNISRLRTEHGVYELPDGSHR-----LLDGMKREPDGSP 1567
 DB 1219 QVAVTPP-----PKETPOISEYQYSELYSASNTSKNWFSELAYLDDLDKL--DDIDE 1272
 QY 1568 YLLAIWTPGET 1578
 DB 1273 YLLNNIMPEKT 1283

RESULT 15
 MAPB_MOUSE STANDARD; PRT; 2464 AA.
 AC P14873;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE MICROTUBULE-ASSOCIATED PROTEIN 1B (MAP1.2) (MAP1(X)) [CONTAINS: MAP1
 DE LIGHT CHAIN LC1].
 GN MAP1B OR MTAP5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=SWISS WEBSTER; TISSUE=Brain;
 RX MEDLINE=90094539; PubMed=2480963;
 RA Noble M., Lewis S.A., Cowan N.J.;
 RT "The microtubule binding domain of microtubule-associated protein
 RT MAP1B contains a repeated sequence motif unrelated to that of MAP2
 RT and tau.";
 RL J. Cell Biol. 109:3367-3376(1989).
 CC -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
 CC -!- PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
 CC THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
 CC TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
 CC MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
 CC STABILIZING MICROTUBULES.
 CC -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
 CC WITH MAP1A AND MAP1B PROTEINS.
 CC -!- DOMAIN: HAS A HIGHLY BASIC REGION WITH MANY COPIES OF THE SEQUENCE
 CC KKEE AND KKEI/V. REPEATED BUT NOT AT FIXED INTERVALS. THIS LATTER
 CC REGION IS RESPONSIBLE FOR THE BINDING OF MAP1B TO MICROTUBULES
 CC BOTH IN VITRO AND IN VIVO.
 CC -!- PTM: LC1 IS COEXPRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED
 CC FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
 CC BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION
 CC OF MAP1B.
 CC -!- SIMILARITY: TO NEURAXIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X51396; CAA35761.1; -;
 CC PIR; S07549; QRMSP1.
 CC MGD; MG1:97179; Mtap5.
 CC InterPro; IPR000102; -;
 CC Pfam; PF00414; MAP1B_Neuraxin; 10.
 CC PROSITE; PS00230; MAP1B_NEURAXIN; 7.
 KW Microtubules; Repeat; Phosphorylation.
 FT CHAIN ? 2464 MAP1 LIGHT CHAIN LC1.
 FT DOMAIN 589 787 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
 FT KKEE AND KKEI/V REPEATS).
 FT DOMAIN 1865 2068 12 X 17 AA TANDEM REPEATS.
 FT REPEAT 1865 1881 1.
 FT REPEAT 1882 1898 2.
 FT REPEAT 1899 1915 3.

FT REPEAT 1916 1932 4.
 FT REPEAT 1933 1949 5.
 FT REPEAT 1950 1966 6.
 FT REPEAT 1967 1983 7.
 FT REPEAT 1984 2000 8.
 FT REPEAT 2001 2017 9.
 FT REPEAT 2018 2034 10.
 FT REPEAT 2035 2051 11.
 FT REPEAT 2052 2068 12.
 SQ SEQUENCE 2464 AA; 270408 MW; FBD3DD99CFDBDA87 CRC64;
 Query Match 2.1%; Score 190; DB 1; Length 2464;
 Best Local Similarity 18.6%; Pred. No. 0.051;
 Matches 291; Conservative 226; Mismatches 640; Indels 410; Gaps 67;
 QY 175 LDSVSQIDKT-----NLGAMNQPLEVSMGNQPKLSTGAKLARDQOQDPLLRNQOCPVP 230
 DB 301 LDRVDSILLTHIGDDNLPGLNSMLQRIAELEERSOGSTNSDMKNLISPDGLGVVFLN 360
 QY 231 ATONTQFPMENOQAWLQKNQOLIGFPEGNQOPMTIRNQOPL-AMGNO-OPMYL---IG 285
 DB 361 VPENLKDPEN-----IRMKRSIEEACFTLQVNLKLSMKPEPLFRSVGNTIEPVILFKMG 416
 QY 286 TPRPALVSGN-----QOLGGPQGNKRPFLFNHQTCLPAGNQL---YGSPTDMH 330
 DB 417 VGKLEMYLNPVKSSKEMQYFMQQTGINKDAELI-----LPGQEVDPISYLSVS 470
 QY 331 QLVNSTGQOQHLLIKNOQP-----SLIRG-----QPCV---PLIDQOATPK 372
 DB 471 SLIYVHPANPAEKIIRVLPFGNSTQYNIIEGLEKHLKHLDFLQPLATQKDLTGQVTPPV 530
 QY 373 GFTHLNMV-----ATS-MSSPGLRPHSQSOVPTTYLHVSVSLRNLTTGTGTCORSR 423
 DB 531 KOVLKORADRESLKPATKPVASKSVKSKETPE-----VVTTSQVEK 576
 QY 424 APAYDSLQDDIHQNKYILSHEISNGCKKALPONSLSLPTPIMAKLEEARG-----KRQ 479
 DB 577 TPKVESKEKVLVKDKPVKTE--SKPSVTEKESVSKKE--QSPVKAEEVAEKQATESKPKVT 633
 QY 480 YHRAMGQTEKHDMLAQIAQSDQVHRNSSTCYEYLDAAKTKIOKVVOENLHGMPEV 539
 DB 634 KQVYVKEIKTKLEKEEKPKPKVKKEDKTLKKDKPKRKEVKEIKK-----684
 QY 540 IEIEDDDTDGAKNKNTASISKAGSKNSSPVKTAKEKICVPTAKKGRAGRKSV 599
 DB 685 -EIKKEKELKKEVKEKTPDKDAKEVKEKKEVKEE--KEPKKELKLSKDKIKSTP 742
 QY 600 -----PPAHASETQLWQPTPKTPLSRSKPKGKR--KSIDSQKARGPSGELLCQDSIA 652
 DB 743 QSDTKKPSALKPKVAKKEESTKKEPLAAGKLKDKGVKVIKKEGKTEAATAVGTAA 802
 QY 653 EIIYRMQNLILGDKEREQEQNAMVLYKGDGALVYESKKRPRKPVDDIDDETRINLLM 712
 DB 803 AAVVAAAGIAASGPVKLEAERSLMSSPEDLTDKDFEELKAE---EIDVAKDIKPOLELI- 858
 QY 713 GKGDEKEDGDEKDKKKEWKEEERRRFRGRADSFIAHMLVQGDRRFRSPKWSGVDSVIG 772
 DB 859 -----EDEKLEKTOP-----GEA-----YVLOKETEVS--KGSAESDEG 892
 QY 773 VFLTONVSDHLSSSAFMSLAARFPFKLSSSRREDNRVRSVVVEDEPGCILNLNIPSQE 832
 DB 893 ITTTEG-----EG-----E 901
 QY 833 KVQHPSDM---EYSGVDSGSKQELRDCNSGIERNFLEKSIONLEEVLSSQDSFDPAI 889
 DB 902 CEQTPPELEPVEKQGVDD-----IEKF-----EDEGAGFEESSETGD 938
 QY 890 FOSGCRVSGSCSCKSDAEFFPTTRCTKTVSGTSVOTGSPNLSDLEICLOGNERPHYEG 949
 DB 939 YEE-----KAETEAEAP-----EEDGEDNAGSASAKHSPTEDDE-SAKAEADVHLKEK 986
 QY 950 -----SGDVQKQETTNVAQKPKDLEKTMNMKDSVCFQGPORNDTNWQTPPSSSYEQCATR- 1003

```

Db 987 RESVVGDDRAEDMDVLEKEAQSEB-----EGEEDKADAREEGEYEPDKTEA 1038
QY 1004 QPHVLIDIED----FGMQGEGELGYSMMSISPRVDRVKNKNVPRFRFQGG--SVPREFTGQ 1057
Db 1039 EDYVMAVADKAAAGVTEQYGYLTSA-----KPGIQSPSREPASS 1081
QY 1058 IIPSPHPELPGMGLSGSSSAVOEHODDHOHNOODENKASHLOKTFLLDLNLSBECLTRQ 1117
Db 1082 IHDET---LPG---GSESEATASDENEDQPEFTATSGYTQSTIEI--SSEPTPMD 1132
QY 1118 SSTKONITDGCCLPRDRTAEDVDPLSNSSNLILVESNNSNKEOTAVEYKETNATILRE 1177
Db 1133 MST-----PR-----DVMSDETNNETEESPOEFVITKYESSLYSQEYSKPAVAS 1178
QY 1178 MKGTLAGKKPTSQWDSLKQVNEGGR--QERNKNMDSIDYEAIRRASISEISEAIKER 1236
Db 1179 FNG-LSEGSK--TDATDG--KDYNASASTISPPSMEEDKFSKALRDAYCSE-----EK 1228
QY 1237 GNMNLAIVRIKDFLERIVKDHGIDLEWLRESPPDKAKDYLLSIRGLGLKSVCEVRLTL 1296
Db 1229 ELKASAELEIKOVSDERLSPAKSPSLSPSPPIEKP-----LGRSVNF--SLTP 1278
QY 1297 HNLAPPVDTNVGRIAVRMGMVPLQPLPESLQLHLLLELYPVLESIOKFLWPRCLKLDORTL 1356
Db 1279 NEIKVSAEGEARSVSGVGTQAVVEHCASPEEKTLEWVSPQSV-----TGSAG 1327
QY 1357 YELHYQLITFGKV--FCTKSRPNCNACMRGCRHFASAYASARLALPAPEERSLTSATI 1414
Db 1328 HTPYQSPTEDEKSSHLPTVSENAQVPYSFE--FSEAKDENERSLSFMDPEVPDSES 1384
QY 1415 PVPPESFPVPAIPIELPLPLEKSLA-----SGAPSNRENCEPIIEEPASPGQECT 1465
Db 1385 PVEKVLPLRSPPLLGSESPYEDFLSADSKVLGRRSSESPFEGKNGKQGPDPRESVSDLT 1444
QY 1466 -----EITESDIEDAYNE-----DPDEIPIKILNI 1491
Db 1445 STGLYQDKOEKSTGFIPIKEDFGPEKKTSDVTWSSQSALALDERKLGCDVSPT--OIDV 1503
QY 1492 EQFGMTLREHMERNMELQEG-----DMSKALVALHPTTT 1525
Db 1504 SQFG-SFKE--DTKMSISEGTVSDKSATPVDEGVAEDTYSHMEGVASVSTASVA---TS 1556
QY 1526 SIPTPKLKNISRLTEHQVYELPDSHRLLDGMDKREDDPPSPYLLAIWTP---GETANSA 1582
Db 1557 SFPEPTDDVS-----PSLHAEVGSPHSTEVDD-SLSVSVVQVQTTTFQETEMS- 1603
QY 1583 QPPEOKC 1589
Db 1604 -PSKEEC 1609

```

Search completed: July 5, 2001, 12:46:19
Job time: 241 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 5, 2001, 12:42:18 ; Search time 36.19 Seconds
(without alignments)
6320.956 Million cell updates/sec

Title: PCT-US01-13059-2

Perfect score: 9089

Sequence: 1 MOSIMSSAVNATEATEQND.....PRPLMARLHPASKLNKNT 1729

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SP TREMBL.16.*

2: SP Archaebacteria.*

3: SP fungi.*

4: SP human.*

5: SP invertebrate.*

6: SP mammal.*

7: SP mhc.*

8: SP organelle.*

9: SP phage.*

10: SP plant.*

11: SP rodent.*

12: SP unclassified.*

13: SP vertebrate.*

14: SP virus.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	3949	43.4	1017	Q9L269	Q9L269 arabidopsis
2	2863	31.5	555	Q9L568	Q9L568 arabidopsis
3	1776.5	19.5	1309	Q9SR66	Q9SR66 arabidopsis
4	1559.5	17.2	1207	Q9SJK6	Q9SJK6 arabidopsis
5	1260	13.9	234	Q9L267	Q9L267 arabidopsis
6	999	11.0	274	Q9FTQ2	Q9FTQ2 oryza sativ
7	709	7.8	917	Q9498	Q9498 arabidopsis
8	281	3.1	2151	Q9NG79	Q9NG79 trichomonas
9	278	3.1	5327	Q97891	Q97891 drosophila
10	245	2.7	2254	Q9HCY0	Q9HCY0 homo sapien
11	241	2.7	2187	Q9H197	Q9H197 homo sapien
12	240.5	2.6	2938	Q61769	Q61769 mus musculus
13	239	2.6	2701	Q9Y520	Q9Y520 homo sapien
14	238	2.6	5120	Q9PU36	Q9PU36 gallus gall
15	235.5	2.6	4880	Q9JLT1	Q9JLT1 rattus norv
16	235.5	2.6	5085	Q9JKS6	Q9JKS6 rattus norv
17	233.5	2.6	2218	Q9SL26	Q9SL26 arabidopsis
18	232	2.6	259	Q9RRQ0	Q9RRQ0 deinococcus
19	232	2.6	2290	Q9VT00	Q9VT00 drosophila

ALIGNMENTS

RESULT 1

Q9L269

ID Q9L269 PRELIMINARY; PRT; 1017 AA.

AC Q9L269;

DT 01-OCT-2000 (TRENBLrel. 15, Created)

DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)

DE HYPOTHETICAL 112.1 KDA PROTEIN.

GN T32M21_160.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

OC Brassicales; Brassicaceae; Arabidopsis.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Bevan M., Terry N., Ardiles W., Buysshaert C., Dasseville R.,

RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,

RA Villarroel R., Gieles J., Van Montagu M., Bancroft I., Mewes H.W.,

RA Rudd S., Lemcke K., Mayer K.F.X.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project.

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

KW EMBL; AL162875; CAB85562.1; -

DR Hypothetical protein.

SQ SEQUENCE 1017 AA; 112139 MW; 4F7C313A891EC83F CRC64;

Query Match

Best Local Similarity 43.4%; Score 3949; DB 10; Length 1017;

Matches 755; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MOSIMSSAVNATEATEQNDGSRQDVLEFDLNKTPQKPSKRRKFKMPKVVVEGPKRRKP 60

Db 235 MOSIMSSAVNATEATEQNDGSRQDVLEFDLNKTPQKPSKRRKFKMPKVVVEGPKRRKP 294

Qy 61 RKAPELPKVVVEGPKRRKRAATQEKVSKETGSAKKKNLKAESATKKPANYGDMNSKP 120

Db 295 RKAPELPKVVVEGPKRRKRAATQEKVSKETGSAKKKNLKAESATKKPANYGDMNSKP 354

QY 121 EVTLKCRKALNFLENPGDARQDSESEIVQNSGANSFSEIRDAIGGTNGSFLDSVSQ 180
Db EVTLKCRKALNFLENPGDARQDSESEIVQNSGANSFSEIRDAIGGTNGSFLDSVSQ 414
QY 181 IDKTNGLGAMNOLPEVSGNQPDKLSTGAKLARQDQDPLLRNQCOPFPVATQNTQPM 240
Db IDKTNGLGAMNOLPEVSGNQPDKLSTGAKLARQDQDPLLRNQCOPFPVATQNTQPM 474
QY 241 NQAWLQWKNOLIGFPFGNQOPRMTIRNQOPCLAMGNOQPMYLGTPRALVSNQOOLGG 300
Db NQAWLQWKNOLIGFPFGNQOPRMTIRNQOPCLAMGNOQPMYLGTPRALVSNQOOLGG 534
QY 301 PQGNKRPFILNHOQCLPAGNOLYGSPTDMHQLVMSTGQOQGLLKNQOPGSLIRGOQPC 360
Db PQGNKRPFILNHOQCLPAGNOLYGSPTDMHQLVMSTGQOQGLLKNQOPGSLIRGOQPC 594
QY 361 VPLIDQOPATPKGFTHLNQMVATSMSPGLRPHSQSQVPTTYLHVESVSRILNGTTCTCQ 420
Db VPLIDQOPATPKGFTHLNQMVATSMSPGLRPHSQSQVPTTYLHVESVSRILNGTTCTCQ 654
QY 421 RSRAPAVDSLOQDHOQNKYILSHSINGNCKKALPQNSLSLPTPIMAKLEEARSGSKROY 480
Db RSRAPAVDSLOQDHOQNKYILSHSINGNCKKALPQNSLSLPTPIMAKLEEARSGSKROY 714
QY 481 HRAMGQTEKHDLNLAQOIAQSDQVERHNSSTCV EYLDAAKTKIKQVQENLHGMPEVI 540
Db HRAMGQTEKHDLNLAQOIAQSDQVERHNSSTCV EYLDAAKTKIKQVQENLHGMPEVI 774
QY 541 EIEDDPTDGARKGNKTASISKGASKGNSPVKTAEEKICIVPKTAKGRAGRKSVPP 600
Db EIEDDPTDGARKGNKTASISKGASKGNSPVKTAEEKICIVPKTAKGRAGRKSVPP 834
QY 601 PAHASEIQLMOPTPKPLSRKPKGGRKSIQDSGRARGPSGELLCCQDSTAEIIRYMON 660
Db PAHASEIQLMOPTPKPLSRKPKGGRKSIQDSGRARGPSGELLCCQDSTAEIIRYMON 894
QY 661 LYLGDKEREQEQNANVLYKGDGALVPYESSKRRKPRKPRVDIDDETRITWNLIMGKDEKEG 720
Db LYLGDKEREQEQNANVLYKGDGALVPYESSKRRKPRKPRVDIDDETRITWNLIMGKDEKEG 954
QY 721 DEEDKDKKKEWEEERRVFRGRADSFARHVLQGD 756
Db DEEDKDKKKEWEEERRVFRGRADSFARHVLQGD 990

RESULT 2

Q9LZ68 PRELIMINARY; PRT; 555 AA.
AC Q9LZ68;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE HYPOTHETICAL 62.1 KDA PROTEIN.
GN T32M21.170.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Terry N., Ardiles W., Buysshaert C., Dasseville R.,
RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
RA Villarroel R., Gleen J., Van Montagu M., Bancroft I., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL162875; CAB85563.1; -;
DR SMART: IPR003265; -;
DR InterPro: SM00478; ENDO3C; 1.

KW Hypothetical protein.
SQ SEQUENCE 555 AA; 62088 MW; A1B44BDDE17FDC1E CRC64;
Query Match 31.5%; Score 2863; DB 10; Length 555;
Best Local Similarity 100.0%; Pred. No. 3.4e-184;
Matches 549; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 785 SSAFMSLAARPPPKLSSSREDENRVSVVVEDPBGCTILNLEIPSWQEKVQHPSDMEVSG 844
Db 7 SSAFMSLAARPPPKLSSSREDENRVSVVVEDPBGCTILNLEIPSWQEKVQHPSDMEVSG 66
QY 845 VDSGSKQLRDCNSGIERENFLEKSIQNLLEEVEISSQDSFDPALFOSCGRVGSCSCSKS 904
Db 67 VDSGSKQLRDCNSGIERENFLEKSIQNLLEEVEISSQDSFDPALFOSCGRVGSCSCSKS 126
QY 905 DAEFPTTRCETKTVSGTSQSVQGTSPNLSDIEICLQGNRPHLYEGSGDVQKQETTNAQK 964
Db 127 DAEFPTTRCETKTVSGTSQSVQGTSPNLSDIEICLQGNRPHLYEGSGDVQKQETTNAQK 186
QY 965 KPDLKTMNKKDSVCFQOPRNDTNWQTPSSSYEOCATRPHVLDIEDFGMOGEGLYSW 1024
Db 187 KPDLKTMNKKDSVCFQOPRNDTNWQTPSSSYEQCATRPHVLDIEDFGMOGEGLYSW 246
QY 1025 MSISPRVDRVKNKVNPRFRQGSVPREFTGQIIPSTHELPGMGLSGSSSAVQEHQDD 1084
Db 247 MSISPRVDRVKNKVNPRFRQGSVPREFTGQIIPSTHELPGMGLSGSSSAVQEHQDD 306
QY 1085 TOHNOQDEMKNASHLQKTFDLNLSSECLTROSSTKQNTDGCCLPRDRTAEDVVDPLSN 1144
Db 307 TOHNOQDEMKNASHLQKTFDLNLSSECLTROSSTKQNTDGCCLPRDRTAEDVVDPLSN 366
QY 1145 NSSLQNTLVESNSNKEQTAVEYKETNATILREMKGTADGKKPTSDMSLRKDXVEGNEG 1204
Db 367 NSSLQNTLVESNSNKEQTAVEYKETNATILREMKGTADGKKPTSDMSLRKDXVEGNEG 426
QY 1205 RQERNKNMDSIDYEAIRRASISEISEIAKERGMNMMNLAIRKDFLERIVKDHGGIDLEW 1264
Db 427 RQERNKNMDSIDYEAIRRASISEISEIAKERGMNMMNLAIRKDFLERIVKDHGGIDLEW 486
QY 1265 LRESPDPKADYLLSIRGLKSEVCRLLTLNHLAPVDVTNIGRIAVRMGWVLPLOLPE 1324
Db 487 LRESPDPKADYLLSIRGLKSEVCRLLTLNHLAPVDVTNIGRIAVRMGWVLPLOLPE 546
QY 1325 SLQLHLEL 1333
Db 547 SLQLHLEL 555
RESULT 3
Q9SR66 PRELIMINARY; PRT; 1309 AA.
AC Q9SR66;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)
DE T22K18.18 PROTEIN.
GN T22K18.18.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC T22K18 genomic sequence.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC010927; AAF04422.1; -;
DR InterPro: IPR003265; -;


```
QY 230 VATONTQPMENQAWLQMKNLQIGFPFGNQQRMTIRNOQPCLAMGNQPMYLGITPRP 289
Db 272 -----QKHWLYSPNL-----QQGM--RYDAIC----- 293
QY 290 ALVSGNOOLGGPOGNKRPIFLNHOTCLPAGNOLYGSPTDMHQLVMSTGGQOGLLTKNQ 349
Db 294 -----SKVFSQQHNVSAF--HATCYSSTSQLSANRV-----LTVEERR 331
QY 350 PGSILRQOCVCLIDQOATP-----KGFTHLNQMVATSMSSPGLRPHSQSOV 398
Db 332 EGIFQGRQSELANVLSKIDTPIKKTTHGHPFNLSMMNKL-----EV 376
QY 399 PTVYLVHVESVRIILNGTGTQCSRAPAYDSLOQDIHOGNKIYLSHSNGCKKALPQ 458
Db 377 PE---HL-----TSGVCSKQO-----QNNKILVDTRVT----- 401
QY 459 NSSLPTPIMAKLEFARGSKRYHRAMQTEKHDLNLAQQAQSQDVVERHNSSTCVELDA 518
Db 402 -----VSKKPKTKSEKQ----- 414
QY 519 AKTKIOKVQENLHGMPEVIEDPTDGARKKNTASISKASKGNSPVKKTAKE 578
Db 415 ---TK-QKNLLPNLCRPPPTGLSPD----- 437
QY 579 KCIVPKTPAKKGRAGRKSVPPPAHASEIQLWQTPPKTPLRSRKPGRKSKIQDSGKA 638
Db 438 -----ELM-----KRNST----- 446
QY 639 RGPSEGLLODSTAETIYRMONLYLGDKEQEONAMVLYK-----GQALVPYE 688
Db 447 -----ETISEL-----LRLLDINREHSETALVPYTMNSQIVLFGGAGAIIVPT 490
QY 689 S-KKRPRPKVDIDDETRINLLMGKDEGEDEKDKKKKEKWEERERVRFRGRADSEI 747
Db 491 PVKPRPKVDLDETRVWKLL--ENINSEGVGSGDEQAKWEEERNVFRGRADSEI 549
QY 748 ARMHLVQGRDRFPKGSVVDSVIGVFLTONVSDHLSSSAFMSLAARFP--PKLSSSREDE 806
Db 550 ARMHLVQGRDRFPKGSVVDSVGVFLTONVSDHLSSSAFMSLASQFPVPFPSSNFD- 608
QY 807 RNVRSVVEDEPGCILNLEIPSQ-----EKVQHPDSMEVSGVDSGSKQLRDCSNS 859
Db 609 -----AGTSSMPTQIYLDSEETMSPDPDHHSSVT----- 640
QY 860 GIERENFLEKSIONLEEVLSODSDFPAIFQSGRVSQSCSKSDAEFFTRCET--KT 917
Db 641 -----LKWTQPDDEKDYVPSNET-----SRSSSEIAISAHESVDKT 676
QY 918 VSQTS--QSQTGSPNLSDBICLQGNRPHLYEGSGDVOK-QETTNVAQKPKDLEKTMNW 974
Db 677 TDSKEYVDSRKG-----SVEVDKTEKCVNLNLFPSDSALTC 716
QY 975 KDSVCFQOPNRDNTWQTPSSSYEQCATRQPHVLDIEDFMQEGGLGYSMWISPRVDRV 1034
Db 717 QHSMVSDAPOT---ERAGSS-----EIDLEGE-YRTSPMKL----- 750
QY 1035 KKNVPRFRFGSGVPRFTGQIIPSTPHELPGMGLSGSSSAVOEHODDTQHNOQDEM 1094
Db 751 -----LQGVQVLSDESNQVSPN-----MSPGDCSSEIKGFQ-----SKKEPT 787
QY 1095 KASHLQKTFDLNLSBE--CLTRQSTTKQNTDGLCLPRDRTAEDVDVDPNLSLQNILV 1153
Db 788 KSS-----VDSSEPGCSQO-----DG-----DVL----- 807
QY 1154 ESNSNKEQTAVYEKTNATILREMGKTLADGKKPTSQWDSLRKDVEGNEGROERNKNNM 1213
Db 808 -----SCQKPT---LKEKGKKVLKE-----EKKAFDWDCLRREAQARAGIREKTRST 852
QY 1214 DSDIYEAIRRAISSEISEAKIKRGMNMLAVRI-----KDFERTVKDHGGIDLE 1263
Db 853 DTDVWKAIRAADVKEVAETIKSRGMNHKLAERTQYLTLANMKINQGGFLDLRLVDHGSIDLE 912
```

```
QY 1264 WLRESPPDKAKDYLLSIRGLGLKSVCEVRLTLHLNLAFFVDTNVGRIAVRMGWVLPQLP 1323
Db 913 WLRDVPPDRAK----- 923
QY 1324 ESLQLHLELLEPVLESIQKFLWPRCLKDORTLYELHYQLITFGKVFCTKSRPNCNACPM 1383
Db 924 -----YELHYQMITEGKVFCTKSKPNCNACPM 950
QY 1384 RGCRRHFASAYASARLALPAPEERSLTSATIPVPPSPFPVPAIMIELPLPLEKSLASGA 1443
Db 951 KGBRHFASAFASARLALPSTKGMGTDPKNPLP-----LHLPFFQREGSEV 999
QY 1444 PSNREN-----CEPIIEEPASPGQCTEITSDIEDAYYNEDPDEIPTKLNIHQFGMT 1497
Db 1000 VQSEPAKKVTCCEPIIEEPASPEPEAEVSIADIEEAF-EDPERIPTRLNDAFTSN 1058
QY 1498 LRHMERNMELQGDMSKALVALHPTTTSIPTPKLKNISRLRTEHQVYELPDSHRLDGM 1557
Db 1059 LKKIMEHNKELQDGNMSSALVALTAETASLPMPLKNISOLRTEHRYVELPDEHPLLAQL 1118
QY 1558 DKREPDDPSYLLAIWTPGETANSAPPEQKCGKSGKCMCFDETCSECNLSREANSQV 1617
Db 1119 EKREPDDPSYLLAIWTPGETANSIOPSVSTCIFIQANGMLCDEETCFSCNSIKETRSQIV 1178
QY 1618 RGTLLI 1623
Db 1179 RGTILV 1184
RESULT 5
QY 09LZ67 PRELIMINARY; PRT; 234 AA.
AC 09LZ67;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HYPOTHETICAL 26.6 KDA PROTEIN.
GN T32M21.180.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Terry N., Ardiles W., Buysshaert C., Dasseville R.,
RA De Clerck R., De Keyser A., Neyt P., Rouze P., Van Den Daele H.,
RA Villarroel R., Gielens J., Van Montagu M., Bancroft I., Mewes H.W.,
RA Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL162875; CAB8564.1; -.
KW Hypothetical protein.
SQ SEQUENCE 234 AA; 26649 MW; B395392E06EB297 CRC64;
```

Query Match 13.9%; Score 1260; DB 10; Length 234;
Best Local Similarity 100.0%; Pred. No. 5.3e-77;
Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1496 MTLREHMERNMELQGDMSKALVALHPTTTSIPTPKLKNISRLRTEHQVYELPDSHRLD 1555
Db 1 MTLREHMERNMELQGDMSKALVALHPTTTSIPTPKLKNISRLRTEHQVYELPDSHRLD 60
QY 1556 GMDKREPDDPSYLLAIWTPGETANSAPPEQKCGKSGKCMCFDETCSECNLSREANSQ 1615
Db 61 GMDKREPDDPSYLLAIWTPGETANSAPPEQKCGKSGKCMCFDETCSECNLSREANSQ 120
QY 1616 TVRGITLLIPCRATMRSFPLNGTYFQVNELFADHSSLPIDVPRDWINDLPRTVYFCT 1675
```


Db 121 TVRGILLIPCRAMRSPPLNGTYFOVNELFADHESLKPIDVPRDWLWDLPRRTVYEGT 180
Qy 1676 SVTSIFRGLSTEQIOFCFWKGFVCGVGEQKTRAPRPLMARLHFPASKLNKNT 1729
Db 181 SVTSIFRGLSTEQIOFCFWKGFVCGVGEQKTRAPRPLMARLHFPASKLNKNT 234

RESULT 6
Q9FTQ2 PRELIMINARY; PRT; 274 AA.
ID Q9FTQ2 AC Q9FTQ2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE P0665D10.14 PROTEIN.
GN P0665D10.14
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
OC Oryza
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
clone:P0665D10.1";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP002861; BAB16489.1; -
SQ SEQUENCE 274 AA; 30951 MW; D7C826BD3982592D CRC64;

Query Match 11.0%; Score 999; DB 10; Length 274;
Best Local Similarity 68.6%; Pred. No. 2.3e-59;
Matches 179; Conservative 38; Mismatches 42; Indels 2; Gaps 2;
Qy 1469 ESDIEDAYNEDDPEIPIKINIEQFGMTLREHME-RNWELOEGDMSKALVALHTTTSI 1527
Db 3 ECAEDSFV-DOPEIPIKINIEFTONLSYMOANNIEEDADMKSALVAITPEVASI 61
Qy 1528 PTPKLNISRLRTEHQVYELPDSPHRLDGMKREPDDPSYLLAIWTPGETANSAPPEQ 1587
Db 62 PTPKLNVSRLRTEHQVYELPDSPHLLGEGFNQREDDPCYLLSIWTFGETAQSDAPKS 121
Qy 1598 KGGKASKMCFDETCSCNSLRANSOTVGTLLIPORTAMRGSPFLNGTYFQVNELFA 1647
Db 122 VCSOENGELCASCTFCSCNSIREAQKVRGTLIPORTAMRGSPFLNGTYFQVNEFA 181
Qy 1648 DHESLKPIDVPRDWLWDLPRRTVYFGTSVTSIFRGLSTEQIOFCFWKGFVCGVGEQKT 1707
Db 182 DHDSRNPIDVPRSWIWLPRRTVYFGTSIPTIFKGLTTEIQHCFWRGFCVGRGDRTS 241
Qy 1708 RAPRPLMARLHFPASKLNK 1728
Db 242 RAPRPLARLHFPASKITRNK 262

RESULT 7
ID O49498 PRELIMINARY; PRT; 917 AA.
AC O49498;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 106.3 KDA PROTEIN.
GN F28A23.180 OR AT4G34060.
OS Arabidopsis thaliana (mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.

RA Bevan M., Weichselgartner M., Fartmann B., Granderath K., Dauner D.,
Herzl A., Neumann S., Hohelsel J., Jesse T., Heijnen L., Vos P.,
Mewes H.W., Mayer K., Schueller C.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Weichselgartner M., Fartmann B., Granderath K., Dauner D., Herzl A.,
Neumann S., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL021961; CAAL7566.1; -
DR EMBL; AL161584; CAB80123.1; -
DR InterPro; IPR003265; -
DR Pfam; PF00730; HhH-GPD; 1.
DR SMART; SM00478; ENDO3c; 1.
KW Hypothetical protein.
SQ SEQUENCE 917 AA; 106310 MW; 302B3770ACE1A454 CRC64;

Query Match 7.8%; Score 709; DB 10; Length 917;
Best Local Similarity 21.4%; Pred. No. 4.1e-39;
Matches 232; Conservative 108; Mismatches 163; Indels 580; Gaps 22;
Qy 651 IAEIYRONLYGDKEREQONAMVLYKGALVPEYSEKRRKPRKVYDIDETTRINWL 710
Db 388 IAKLIKMDGRLLKINKK-----VTTMIKADKKLV-----TAKVNLDPETIKEDV 431
Qy 711 LMKGDEKEGDEEDKKKEKWEERVRFRGRADSFIAHMLVQGDRESPKSGVDSV 770
Db 432 LM--VNDSPSRSYDDKETEAKWKEREIFQTRIDLFINRMHRLQGNRKPKQKSGVDSV 489
Qy 771 IGVFTQVNSHLSSAFSMLAARPPKLSRREDRNVRVSVVVEDPEGCILNLEIPSW 830
Db 490 VGVFTQNTFTYLSNAPFSAVAKEP-----VDARGLSYVIEEPQ----- 530
Qy 831 QEKVOHPDSMEVSGVDSGSKQELRDCSNSGIERFNLEKSIQNLSEEVLSODSFDAIF 890
Db 531 -----DAKSSCCI-----ILSDE----- 538
Qy 891 QSCRGVSCSKSDAEFTTTCETKTVSGTSQSVQTSPLNSDSEICLOGNERPHLYEGS 950
Db 539 -----IIII----- 543
Qy 951 GDVQKQETTNVAQKPDLEKTMNWKDSVCFGQPRNDTNWQTTTPSSSYEQCATRQPHVLDI 1010
Db 544 ----- 543
Qy 1011 EDFGQGEGLGYSMWSISPRVDRVKNKNVPRFRFGGSPREFTGQIIPSTPHELPGMG 1070
Db 544 ----- 543
Qy 1071 LSGSSSAVQEHODDTQHNOQDEMKNKASHLQKTFDLLNSSECLTRQSTKONTGCLP 1130
Db 544 ---SISKVEDHEN-----TAKRKNKXTGII- 565
Qy 1131 RDTAEDVVDPLSNNSLQNLIVESNSSNKEQTAVEYKETNATILREMKGTLDCKKPTS 1190
Db 566 ----EDEIVD----- 571
Qy 1191 QWDSLRKQVEGNEGRKRNKNMBSIDYEAIRRASISEISEIAKERGMNMLAVRIKDFL 1250
Db 572 -WNLLR-MYTKEG--SRPEHMDSVNMSDYLRSQNVLETITKKRGQFRILS----- 620
Qy 1251 ERIVKHGGIDLEMLRSPDPKADKYLISIRGLGLKSECVRLLTFLHNLAFVDPNVGRI 1310
Db 621 ERIL----- 632
Qy 1311 AYRMGWVPLQPLPSLQLHLELPVLESIOKFLWPLCKLDQRTLYELHVLQITFGKVF 1370
Db 633 AVRLGLVLEPLPVGQMH-----QLFYEYELHYOMITFGKVF 669


```
Db 1500 SOTESKPNLSLLPPKQYL-GLKLTSSNGLKLPKLPNRRNONQAKVEKET- 1554
QY 1175 LREMGKTLADGKKPTQWDSLRKDVGNRGQE- RNNKNMDSIDYEAIRRAISSETS 1230
Db 1555 -----KEYKAEPKELGFKSDFKIEGASFKGINLSSGKPF- 1591
QY 1231 EAIKRGNNMLAVRIKPLEIRIVKHGDIGIDLEWRRESPDKA-KDYLL-----SI 1280
Db 1592 EYPSGKSFDGMIKQVKNKAEENKPKFSEFNLNLNANKTPLSNAPKPIALSNDGKKPSSL 1651
QY 1281 RGLGLKSVCEVRLTLHLNLAFFVDNMGRIAVRMGMVPLQPLPESLQHLHLELYPVLES 1340
Db 1652 AGLNISKSSLASVSINALDKSGD-DVPKLAUSIDKLRKEPPPP-----PTPPAV 1700
QY 1341 QKFLWPLRC- KLDQRTLYELHYLIITGKVFCTKSRPNCNAC-PMRGECRHFFASAYAS 1396
Db 1701 TK- PALATEKEIRESSIADV-TPLAMITKINSKAK-KPDFSALKPNLGNKQSSSSNEQ 1755
QY 1397 ARLALPAPERSLTSATIPVPPESSPPVAIPMIEPLPLEKSLAG-APSRENC 1450
Db 1756 KPLNPNFSALKLPQKSEETKPNLSSLLPPKQTLGLKLPKLTSSNGLKLPKLPNQNQ 1815
QY 1451 EPIIEE-----PASPGQECTEITESDIEDAYN-----EDPD----- 1482
Db 1816 NOEVKESKEVKAKEPKELGKSDFKIEGASFKGINLSSSKGFEVPSGKSFDDMIKQKA 1875
QY 1483 --EIPITKILNIQFGMTLREHMERNEMLQ-----EGMSKALVALHPTTTSIPTP 1530
Db 1876 KTEQPKTEBEPKFFNLNLNANKTPLSNAPKPTALSLSNDSNKGQVOCALITIRPIP-- 1933
QY 1531 KUKNISRLRTEHQVVELPDSHLLQMDKREDDPSYLLAIWTPGETANSQAPPEQKCG 1590
Db 1934 -LSKGDKPRADQSLKLGKS---LSNPTQNKPD-----LSSLKPPQKSEE 1974
QY 1591 GKAS 1594
Db 1975 NKPS 1978

RESULT 9
ID 076891 PRELIMINARY; PRT: 5327 AA.
AC 076891;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE EG:49E4.1 PROTEIN.
GN FUTSCH OR EG:49E4.1 OR CG3064.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RN SEQUENCE FROM N.A.
RP Papagiannakis G., Spanos L., Siden-Kiamos I., Louis C.;
RA "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RP Benos P.;
RA Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RL EMBL; AL031128; CAA20006.1; -.
DR FlyBase; FBgn0025392; futsch.
SQ SEQUENCE 5327 AA; 575942 MW; FEFEE23A118FF38A CRC64;

Query Match 3.1%; Score 278; DB 5; Length 5327;
Best Local Similarity 19.7%; Pred. No. 4.8e-09;
Matches 277; Conservative 198; Mismatches 583; Indels 348; Gaps 59;

QY 2 QSIMDSSAVNATEAQDNGSRQDVLFDLNTPOQPSKRRKRMKPVVVEG-----K 55
```

```
Db 1986 ESIKD-----EAESKESRRE-----SVAESPLPSKEASRPASVAESIKDEAK 2031
QY 56 PRKRPRKPAELPKVVGKPKRP-----RKAATQEVKSKETSAKKNNLKESATKK 108
Db 2032 SKESRRESVAESKPLPSKEASRPASVAESIKDEAKESRRESAEAKSPLPSKEASR 2091
QY 109 PANVG-----DMNKSPVTLKSCRKALNFDLENPGDAR---QGDSE---SEIVQNSSGANSF 160
Db 2092 PASVAESVDEADKESRRE-----SMAESGKAQSIKGDQSPLEKVSREPESVAESV 2144
QY 161 SEIRDAIGTNGSFLDSVSQIDKTNCLGAMNOPLEVSMGNQPKLSTGAKLARDQDOLL 220
Db 2145 KD--DPVKESKESRRESVAGSVTADSDQSPLESKASRPESVVDVSKDEAKESR 2202
QY 221 TRNQOQFPVATQNTQFPMENQOQWLQMKNLQIGFPNQOQPRMTIRNQOQCLAMNQOP 280
Db 2203 ESKTESVIPKAKDDKSPKEVLQ-----PVSMT-----TIREDA-----DQP 2240
QY 281 MYLIGTPRALVSGNQ-----LGGPGNKRPIFLNHOTCLPAGNQLYGPTDMHQLV 333
Db 2241 M-----KPSAERRESRETAESIKASSPRDEKSPLASKEASRPGSVAESIKYDLDPQII 2294
QY 334 MSTGGQOH-----GLLIKN-----QQPGSLIRGOQPCVPLID---QQPATPK-----GF 374
Db 2295 KDKSTEHSRRESLEDKSAVTSEKSVSRPLSVASDHEAAVAIEDDAKSISPKDKSRPGF 2354
QY 375 THLNQVATSMSPGLRPHRSQSQVPTTYLHVESVR-----ILNGTTGTQCSRAPAYD 428
Db 2355 -----VAETVSP-----IEATMEFSKIEVVEKSSLSLSLOGSGGKLQDSSPV-- 2400
QY 429 SLOODTHOQNKYLSHEISNGNGCKKALPONSPLPTPIMAKLEEAARGSKRQYHRAMQOTE 488
Db 2401 ---DVAEGD---FSAVA-----SVSTVPTLTTPAELA-----QIGAATV 2436
QY 489 KHDNLNLAQIAOS-QDVERHNSSTCVEYLDAAKTKTKIOKVQENLHGMPEVTEIEDD-- 545
Db 2437 SSPLEALRTPSAPEHISRADSPAECASEIASQDKSPQVLKES---SRPAWVAESKDDAA 2494
QY 546 ---PTDGARKGKNTASISKAGKN---SSPVKKT-----AEKCKIVPKTPAKKGRA 592
Db 2495 QLKSSVEDLRSVPASTEISRPSASAGETASSPIEEAPKDFAEFEQAEKAVLPLTELKGNL 2554
QY 593 GRKKSVPVPAHASEIQLWOPTPPKPLSRKPKGKRKSIQDSGKGARGSGELLQDSTA 652
Db 2555 PTLSSPVDVAHAS-VQPAELSKVDIEKTASSPIDEASPIDEASPAERPEPESPAESAKDAA 2613
QY 653 EIIYRMQNLVLDGKEREQEQNAWL-----YKGDGALVPVYESKKRKPVKVDIDETT 705
Db 2614 ESV-----EKSADSRPSPSVVESTKADSTKGDISPSP-ESVLEGPADDVEKSKES 2663
QY 706 RIWNLIMGKDEKEGDEEDKDKKKEKWEERVFRGRADSFIAHMLVQGRFSPWKS 765
Db 2664 RPPSV---SASITGDSKDSR-----PAS 2685
QY 766 VYDSVIGFELTONVSDHLSSAFMSLAARPPKLSRREDERNVRSVVVEDEPGCILNLN 825
Db 2686 VYESV-----KDEH-----DKAESRRESIAKVESVI--DEAG----- 2715
QY 826 EIPSMQEKVQHPSDMEVSGVDSGSKSEOLRDCSNSGIEREFNFEKSIQNL-----EVL 879
Db 2716 -----KSDSKSSQDSQKDEKSTLASKEASRRESVSVESKDDAEKESRPSVI 2764
QY 880 SS-----QDSFDPALFQSGRGVSC-----SCSKSDAEFPPTTRCTKTVSGTS 922
Db 2765 ASGEVPVRESKPLDSKDTSRPGSVVESVTADEKSEQSRRESVAESVKAOTKK-DGKS 2823
QY 923 QSVQGTSPNLSDEICLOGNE-----RPHLYEGSDVQKQETTNVAOKKPLEKTMNKDVS 978
Db 2824 Q--EASRPSSVDELLKDDDEKQESRRQSIITGSHKAMSTMGDESPMDKADKSEPSRPSV 2881
QY 979 CFGQPRNDTNWQTTSSSVEQCATRQPHVLD-IEDFGMQEGELGYSWMSIS-PR--VDRV 1034
```

```
Db 2882 AESIKHENTKDEESPLGS-----RRDSVAESIKSDITKGEKPLSPSKEVSRPESVVGSI 2935
Qy 1035 KKNVPRFRFGGSGVPREFTGII---PSTPHLPGLMGLSGSSSAVQEHQDDTHQHQD 1091
Db 2936 KDEAESRRRESVAESPESKDATSAPPSKEHSRP-----ESVLGSLKD 2980
Qy 1092 EMKASHLQKTFLLNLSSECLTRQSGSTK-----QNTDGLCPDRDRTA-----EDVVD 1140
Db 2981 EGDKTTSRVSVADSIKDEKSLLSVQSEAPSEAESESLKDAAPSOFTSRPESVTSVKD 3040
Qy 1141 PLSNSSLQNLIVESNSNKEQTAVYKFNATILREMK-GTLADGKKPKTSQWDSLRKDV 1199
Db 3041 GKSPVASKASRPASVAENAKDSADESQEPESLPQSKAGSINDEKSLASKDAEAKSK 3100
Qy 1200 EGNREGROERNKNNMDSIDYAIRASISEI--SEAIRKGMNNML-----AVRIK 1247
Db 3101 E--ESRRESVAEPPLVSKEVSRPASVAESVKDEAKSEESPLMSKEASRPASVAGSVK 3158
Qy 1248 DFLERIYKDHGGIDLEWLRESPPDKA 1273
Db 3159 DEAEKSKE-----ESRRESVAESK 3177

RESULT 10
Q9HCY0
ID Q9HCY0 PRELIMINARY; PRT; 2254 AA.
AC Q9HCY0;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE PUTATIVE TRANSCRIPTION FACTOR-LIKE NUCLEAR REGULATOR.
GN TFNR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RA SEQUENCE FROM N.A.
RP Kelter A.R., Herchenbach J., Wirth B.;
RT "The transcription factor like nuclear regulator (TFNR) contains a
RT novel 55-amino acid motif repeated 9 times and maps closely to SMNL.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ238520; CAC04245.1; -.
SQ SEQUENCE 2254 AA; 252844 MW; F350E96F3F04CFE CRC64;

Query Match 2.7%; Score 245; DB 4; Length 2254;
Best Local Similarity 18.7%; Pred. No. 2.3e-07;
Matches 358; Conservative 250; Mismatches 722; Indels 584; Gaps 85;

Qy 11 NATEATEONDGSRD-----VLEFDLNTPOOK---PSKRRKFKMPKVYVGGKPKRK 59
Db 339 NKFKREKTNGRIDKAFQEKRPDFDFFAHLQKVLAEBEKKQKSVKNHLSKEKSKTK 398
Qy 60 KRPAPLFPVVVEG-----KPKRKPRAAT----- 84
Db 399 PRKNVKKVYACGVNNDPDESMSRISDTSQKDAQTVEESLTLSDREDAEQVALEVD 458
Qy 85 --QKVKSKEGSAK---KNLKESATKK--PANVGDMNK-----SPEVTLKSKRK----- 129
Db 459 LNKRRRRKQDQGANELGVNLLNENATVQAGPSGKHKHKNQCAIRPELKEGECSEKQML 518
Qy 130 --ALNED-----LENPGDAQDSESEIVQNSS-----GANSFS 161
Db 519 SCTQNDIGVGFASTEKEVKEKRTDPILSLQDQATSVATSESSSSTDLPSFEVIGALC 578
Qy 162 EIRDAIGG----TNGSFLDSVQIDKTNGL-----GAMNQPLEVSMGNQDKLSTGAKLA 212
Db 579 EVNNAEGSCIEERNVDLKNLSLEIDQENVKPMLRGFRQP-----KPNLSRAGKSV 631
Qy 213 RDQ--QPDLLTRNOCCQFPVATQNTQTFPMENQQAQLQKNKQLIGFPFGNQPFRTIRNQOP 271
Db 632 LSOGKTESEKNSKSHKTSVEKNHVEKDKMNTLDILRMETTERENP-----EAEVTSVLGEKN 688
```

```
Qy 272 CLAMGNOOPWYLIGTPRALVSGNOQLGGPOGNK-----RPIFLN-----HOTC 315
Db 689 CLOEGSQ-----LKALRPVQVRGLQPKPNAGKAERKEILISQEEIGANVEKNENESC 743
Qy 316 LPAGNQLYSPTDMHOLVMTGGQQHGLLIKNOQPGSLIRGQQP-----CVPLIDQQ 367
Db 744 A-----DRDTQPMHEDQSRKDFEEDVILQPEKNDQFQNVQDEPKVLNLSVQENN 796
Qy 368 PATPKGFTHLNOMVATSMSSPGLRPHSQSOVPTYLHVESVSRLLNGTTGTCQSRAPAY 427
Db 797 KA-----NKLNOV-----PILTRFRQPKP-----NIGRTGRREISKE 831
Qy 428 DLSQDDTHOQNKYILSHEISNGNGCKKALPONSSLPTPIPAKLEARGSKRYIRAGQOT 487
Db 832 EVLEKILVSEMAAALRETVLDTSPKEM-----VPAINTK-EMQSDLKETGRAISPR 885
Qy 488 EKHDNLNAAQIAQSQDVERHNSSTCV-----EYLDAAKTKIKQKVQV-----NLHG 534
Db 886 EKILDVIDDTEMETGLKAMGREICLREKTPVIDATE--EIDKLEEAGRREISPOKNG 943
Qy 535 MPPEVIEIEDDPTDGARKGKNTASISKGASGNSSPVKKTAEKCI--VPKTPAKGR- 591
Db 944 -PEEVKPLGEVETD-----LKATGNSSPREKTPGEGTDAEEDKNEETGRR 990
Qy 592 -----AGRKKS-----VPPPAHAS-EIQLWQP-----TPPKTP 618
Db 991 KISPRENGPEVXPVDEMETHLNTATGREGSPREKTPVIDATEIDLEETEREVSPQENG 1050
Qy 619 LRSKPKGKGRKSIQDSGK---ARGSGELLQCQDSIAEIIYRMONLYLGDKERE--OEQN 673
Db 1051 LEEVKPLGEMETDLKATGRDFFRGKTPVI--DAIEI-----EIDLEETEREISPOEN 1103
Qy 674 AMVLYKGDGAL---VPYESKRRKPRKPV-DIDDETTIWNLLMCKGDEKDEKDKKE 729
Db 1104 GLEEVKPLGEMQTLKATGREISPREKTPVIDAT-----EIDKLE 1146
Qy 730 KWEERERRVFRGRADSFIAHMLVQGRDRFSPWKSGVVDVIGVFLTQNVSDHLSASF 789
Db 1147 E-----TGREISP-----EENGPEEVKPVDEM 1169
Qy 790 SLAARFPKPLSSREDERNVSVVVEDEPECILNLIPIPSQEKVQHPDSMEVSGVDSGS 849
Db 1170 ETDLKTTGREGSGREKTR-----VIDAAEVETDLEE-----TEREISPOENGP 1214
Qy 850 KEQLRDCSNGIERFNFLEKSIQNLLEEVLSSQDSFDPALFQSCGRVGS CSCSKSDAEFP 909
Db 1215 EE-----VKPVGKMETDLKEIREEISORE-----KVLAEFS 1245
Qy 910 TTR-----CET-----KTVSGTSQSV-----QTGSPNL-----SDEICLOGNER 943
Db 1246 AIREKEIDLKETGRDIPIMEKYSGKMAVVEEMADLKETGKFNFRGSEECV----- 1300
Qy 944 PHLYEGSDYVQKQETTINVAOKPDLKTMNWKVSCVCGQPRNDTNWOTTPSSSYEQCATR 1003
Db 1301 --TEEKVAELKQTKTDISPENELETSTR-----OTDTHLMQSGNDPSAMPS- 1349
Qy 1004 QPHVLIEDIEFGMOGEGLYGWSNISPRVDRVKNKNVPRP--FRQGG----SVPREFTGQ 1057
Db 1350 ----LDIQNTSSE-----VLSMMHTPVVEERKNESEKSHSFKISSOTHESEKTEVOGI 1401
Qy 1058 IIPSTHELPGMGLSGSSSAVQEHQDDTHQNDQEMKASHLQKTFLLNLSSECLTRQ 1117
Db 1402 QSPDVPEQFSDINLSKSL-----POEQKPLEIKPAPFVRSRF----- 1438
Qy 1118 SSTQNITDGLCPDRFTAED--VVDPLSNSSSLQNLIVESNSNKEQTAVYKFNATIL 1175
Db 1439 KRKPNLARAALKRETTSESEKIYVEKKSETKMETIVQENNEQDTLPLSQHDEASLMS 1498
Qy 1176 REMKGTILADGKK-----PTSQW-----DSLKRDVEGNEGROERNKNNMDSI 1216
Db 1499 RE-KDTLIGHRNEEAVILPCTQTERNLSPNSCPKEESQAPVQKNDVSVVSGTNNVNTF 1557
```


Query Match 2.6%; Score 238; DB 13; Length 5120;
Best Local Similarity 18.3%; Pred. No. 2.2e-06;
Matches 379; Conservative 247; Mismatches 689; Indels 760; Gaps 91;

QY 7 SSAVNATEQNDGSRQDVLFDLNLKTPQOKPSKRKRKFMPKVVVEGKPKRKP 62
DB 229 SQAPPTKPSLOQSGS-----VKQSPQAPKQGGVPRKPSAQAGPPKQOPGSEKPT 279
QY 63 -----PAELPKVVVEGK-----PKRKRKAATQ-----EKVKS 90
DB 280 AQOTGPAKOPPGPGKTPLOQTGPVKVPPQAGPTKFSQOTAGAKSILAQQGLTKPKG 339
QY 91 KTGSAKKNLKESATKP-----ANVGDM-----116
DB 340 QOPGPEKLPQOQASTQVSTPKTKFCPLCTTLLHTPEKANYTCTOCHTVVCSL 399
QY 117 ---NKSPVT-----LKSC---RKALNFDLEN-----137
DB 400 CGFNPHITEKEWLCNLCQMRALGDLGSHGGPGLPPPKQKTPPTASTAKPSQL 459
QY 138 -----PGDAGQDSESEIVONS-----SGANSFSEIRDAIGGTNGSFELD 176
DB 460 QPGQKDKASPKPDPQADSKKPPVKQKPSMPGSPVPKSKQTHAEPD-----TGQOID 514
QY 177 SV---SQDKTNGLAMNQ---LEVSMGNQDPKLTGAK---LARDQDDLLTRNQOQFP 229
DB 515 STPKSDOVKPTQAEKQKQPSIQKPTMDTPTTSAAPGVKQDLADPQSPS---TQOKVTDSP 572
QY 230 VATQNTQPPMENQ-----QAWLQMNQILGFPFGNQPPRMTRNQOQCLAMGN 277
DB 573 M-PETTKPPADTHPAGDKPDKLPQVSRQKSDPKLASOSGAKSDAKTKQKSEPAVPKDD 631
QY 278 QOPMYLIGTPRALVSGNQLGPGQGNRPFLNHTCLPAGNQLYGSPTDMHQLVMSTG 337
DB 632 PKLQTKPAPKPTKPAK---GPOAGTGPRTSAQA-PQOQPKTPQOSRRSLNLG 687
QY 338 G-----QOGLLNNQPGSLIRGQOQCVPLIDQOP 368
DB 688 GITDAPKQPTTQETVTKGLFGFASIFSQASSLIATAGPGSQTSGPAPPATKQOPP 747
QY 369 ATPKGFTHLQNVATSMSSPGLRPHSQOVPTTYLHVESVRILNGTGTTCORSRAPAYD 428
DB 748 SOPPASQAPPEAAQAPPPKAAPTKKTETP---LASEKLGPMASDSTLTTKGSDLEKKP 804
QY 429 SIQODI-HQGNKYILSHSINGNCKKALPQNSLSUPTTMAKLEEARSGKQYHRAMGT 487
DB 805 SLAKDSKHQTAEKAPAELEQ---EKASQPKVSCPLC-----839
QY 488 EKHDNLQAQIAQODVHRNSSTCVELDAAKTKIQVQVENLHGM--PPEVIEIED- 544
DB 840 -KTGLNIGSK-----DPPNFNTC-----TECKKVVVC-NLCGFNPMPHIVEVQEW 881
QY 545 ---DPTDGARKG---NTASISGASKGNSPVKKTAENEKCIVPKTPAKKGRAGKK 596
DB 882 LCLNCQOTRAMSGQLDMGKVPKLPK---LGSPQVSKP-----PATP-----QKQ 923
QY 597 SVPPPAHASEIQLOMPTPKPTPLSRSKP-----GGRKSI-----632
DB 924 PVPVSHSPKSSSTPTTAAT-----KKEEVPKVPKPLQOGLKLETLSDAKTQOQIG 978
QY 633 -ODSKARGPSGELCQSDIAEIIYRMONLYGDKEREQEQNAMVLYKGDGALVPYESSK 691
DB 979 KEDAKSKQGLFKFTPSADKIQRVSKQEDSRLOQTCLTKTPSSDKILHGQVEDIKFQEA 1038
QY 692 RKPVRKVD-----IDDETRTNLLMGKDEKEGDEEDKKEKKEWEEERRVFRGRADSF 746
DB 1039 LAKTPSADKILHRQKQEDPKLQOMKMAKALSADKIQPEAQKEDVOLQEVRLSKAVSADKI 1098
QY 747 IARMLVGDRDRFSPWKSVDVSVIGVFLTPQNVSDHLSSSAFMSLAA-RFPKLSRSSRED 805
DB 1099 ---QHGIQKDLNLQHVKTEKTSVSEKIQEAQESKLOQDKLPKTLSEDKIPATVSS---1152

QY 806 ERNVRSVVVEDEPGCILLNLNEIPSMQEKVQHPSMEVSGVDSGSKQLRDCNSGIERRN 865
DB 1153 HKKLLSKSEEDKKPELL-----EKSTPHPKD-----KKEQI-----1183
QY 866 FLEKSIQNLLEEVLSSQDSFDPALFQSGRGVSCSCSKSDAEFPTRCETKTVSG--TSQ 923
DB 1184 -----TAETTHITQ 1194
QY 924 SVQTSPLNSDEICLQGNRPHLYEGSDVQKOE--TNVAAKPKPDLKTMNKDVSVCFO 982
DB 1195 KVEVEAP-----CDKLHEKK-----QEDVKKEDLTGTGPQVSKPEKAEKTPVPSR 1243
QY 983 -PRND-----TNWQITPSSSYEQCATRQPHVLDIEDFGHQGEGLGTSWMSISPR 1030
DB 1244 LPRSDHVEAVREKIEKEDDKSDTSSQOQ---KSPQ---GLSDTGYSSDGISSL-----1292
QY 1031 VDRVKNKVPFRFRQGGVPRFTGQIIPSTPHPELPGMLSGSSSAVQEHQDDQHNOQ 1090
DB 1293 -----GEIPSH-----IPSEKDL-----REPSQK 1313
QY 1091 DEMNKAS-----HLQKTFLLDNLNSBECLTROSSTKQNTDGLCLPRDRTAEDVDP 1142
DB 1314 DTISOESPSPDLAKLESTVLSLEAQASTLTDEKSVK-----KELYET 1360
QY 1143 SNSSLON-----TLVPSNNSNKEQTAVYEKETNATILREMKGTADGKKKTSQWDSL 1196
DB 1361 SEQTKDOHKTKPLPTPESYSSEEDLEA-IOEGERTIAADSKG-----GASSQTDYKE 1413
QY 1197 KQVEGNE--GROERNKNMDSIDYE-----AIRASI-SEISEAIK-----ERGMNMLAV 1244
DB 1414 ED-GGNDTPARKQRVDSVEDSESESPVPRKRKRASVSGSSSDEYKRRDDSGSGDE---1469
QY 1245 RIKDFLERIVKD-----HGGIDLEWL-----ESPPOKADYLLSIRGL 1283
DB 1470 --EDFIRKQIEMSADASGSEDDDEFIRNQLKEISVTEOKKEEVKSKAKGTGVKHRM 1527
QY 1284 GLKSEVCEVRLTLHLNLPVDTNVGRIAVRMGW-----VPLQPLP-----ESL 1326
DB 1528 ARKS-----SAGYDEADAGR--RHSWHDDDDTDFDESPEPKYRETKSQDGEEL 1572
QY 1327 -----QLHLELVPVLESIQNLFMPRLCKLQORTLY-----ELHYQLTLFGKVCT 1372
DB 1573 AISGGGLRRFKTIELNSTITS-----KYSETPEQKGLIYFDEEPELEMESLT-----1621
QY 1373 KSRPNCNACPMRGEGRHFASAYASARLAPAEERSLTSATIPVPPESPVPVPIPIELP 1432
DB 1622 -----DSPDRSGSGSSSLHASSFTPGTSPTSVSS 1652
QY 1433 L-----PLEKSLASGAPSRENCE-----PIIEEPASQGECTEITESDIEDAYN 1478
DB 1653 LDESDSSPSHKLGGESKQQRKARHSGPLLTIED--SSEELREEEELKKEQKOR 1711
QY 1479 E-----DPDEI-----PTIKNIEQFGMTLREHMERNMELQEG 1511
DB 1712 ELEQOQRSSSKSKKDKDELRAQRRRERKPTPSNLSPIEDASPT--EELQAAEMEE- 1768
QY 1512 DMSKALVALHPHTTTSIPTPKLKN-----IS---RLRTEHQVVELPDSHRL-----LDG 1556
DB 1769 -----LHRSSCEYSFSTIESDEPGEISEPEKIEIEQVKVYKLTAVSLYSPTEKLG 1820
QY 1557 MDKRPDDPSYLLAIWTFGETANSAPPEQKCGKAKSGKMCFTDCTECNLSREANSOT 1616
DB 1821 ALKEESGQ-----KTLKSAEEVYEMIKHTKSKSF-QIASEKDEVFE--RES 1865
QY 1617 VRGTLIIICPTAMRGSFPLNGTYFQVNFELFADHESLAPIDVPRDWINDLPRTYVFGTS 1676
DB 1866 LYGGMLI-----EDYIYESLIEDTYNGTV 1889
QY 1677 VTSIF--RGLSTEQIQFCFWKGVCVGRFEQKTRA 1709
DB 1890 DTNLMRODESNEYIQ-----QKQKKEKIRA 1915

RESULT 15
Q9JLT1
ID Q9JLT1 PRELIMINARY; PRT; 4880 AA.
AC Q9JLT1;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE MULTIDOMAIN PRESYNAPTIC CYTOMATRIX PROTEIN PICCOLO.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20170257; PubMed=10707984;
RA Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,
RA Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;
RT "Piccolo, a Presynaptic Zinc Finger Protein Structurally Related to
RT Bassoon";
RL Neuron 25:203-214(2000).
DR EMBL; AF138789; AA07822.2; -.
DR InterPro; IPR000008; -.
DR InterPro; IPR001478; -.
DR InterPro; IPR002965; -.
DR Pfam; PF00168; C2; 1.
DR PRINTS; PR01217; PRICHXTENS.
DR PROSITE; PS00499; C2_DOMAIN_1; UNKNOWN_1.
DR PROSITE; PS00004; C2_DOMAIN_2; 1.
DR SMART; SM00239; C2; 1.
KW Matrix protein.
SQ SEQUENCE 4880 AA; 530148 MW; 73951BE4ED83EA68 CRC64;

Query Match 2.68; Score 235.5; DB 11; Length 4880;
Best Local Similarity 19.38; Pred. No. 3e-06;
Matches 362; Conservative 239; Mismatches 666; Indels 609; Gaps 93;

QY 35 POQPSKRK---RKMPKVVEGKPKRKPAPLPELVVVEGKPKRKAATQEKVKS 90
DB 612 PSQASPKKEPPSKDSFKALESKKPP-EPKPPPE-PKKPPEK---KPPPLVKQPTLHG 666
QY 91 KETGSAKKNLKE-----SATKKP-----ANVGDMSNKSPVTLKSCRKALNFDLE 136
DB 667 PTPATAPQLPVAELPEAPPEPPSGPLPEQAKAPVGDVEPKPKMT-----ETRADIQ 720
QY 137 NPGDARGDSESEIVQNSSGANSFSETRDAIGTNGSFLDSVSGIDXTNGLGAMNQPLEV 196
DB 721 SSSTTKP-DILSSQVQSAQVKTASPLK-----TDSAKPSQSFPPTGKTTPLDS 769
QY 197 SMGNOPDKLSTGAKLARDQPDLLTRNOQCQFPVATONTQPMENOQAWLQMKNLIGFP 256
DB 770 KAMPRP---ASDKIIISQPGSGESKDPKHIDPI--QKKDEPKKAQ-----P 811
QY 257 FGNQOQRMTRNOQPCILAMNQOQPMYLLIGTPRA---LVSGNQOLGQPGQGNKRPFLN-- 311
DB 812 KGSFKP-----ETKVPVKSGTPS---GT-RPTAGGAAPPQQPPKPPQSRRESNLNG 861
QY 312 -----HQTCLP---AGNOLYG---SPTDMHQLVMSTGGQGHLLIKNQOQSLIRGQQ 358
DB 862 GITDAPKSPQTTTQETVTGTLFGFASIFSOASNLISTAGQ-----GPH 906
QY 359 PCVPLIDQOPATPKGFTHLNQMVATSSSPLRPHSQS---QVPTYLHVSVSRILGTT 416
DB 907 P-----QTGPAAP-----SKQAPTQSQAAGPAKSTQCLPPAPAKATAVKEAKAAA 955
QY 417 GTCQSR---APAYDSLQDDIHOQNYILSHSISNGCKKALPQNSSLPTPIAKLEEA 473
DB 956 AENLESPEQAPTAKTEKD-----KKPPPAKVGVKPPP--SEPEKA 994
QY 474 RGSKRQYHRAMGOTE-----KHDNLNLAQQIAQSQODVERHNSSTCVYLDRAKTKIQ 525
DB 995 VPA-----HKPDKTKPKPACPLCRTELNLGSG-----EPPNFNTCTEC-----KNQVC 1038

QY 526 KVVQENLHGM--PPEVIEIED-----DPTDGARKGK--NTASISKGASKGNSSPVKKTAE 576
DB 1039 -----NLCGFNPPTHLTEIOEWLCLNCQTORAISGQLGDMKMPAPSPGKASWPAPAE 1093
QY 577 KECIYVPTP-----AKGRAGRKKSVPPPAHASEIQLMQPTTPKPLSRSPKPKGKGR 630
DB 1094 PSS---QKTPTGTQVKGKKEAEKTEAEKPVPEKETASTIEKTPPMVTTDQKLEESGKK 1150
QY 631 SIODSKARGPSGELLQCDSIAELIYRMQNLVGLDKREQEONAMVLYKGDGALVPYESK 690
DB 1151 SKVSA-----LPEKKPSSEKA-----ISADKK 1173
QY 691 KRKP-----RPKVDIDDETTRIWNLLMGKDEKE----- 719
DB 1174 ERAPPAEKEPPEKRPDPVKKLPPPEAKPLSSEGEKHEILKAHVQIPPEEPTGKVAAK 1233
QY 720 -GDEEK--DKKKE-----KMWEEERRVFRGRADSFIAHMLHVQDGRFRSPWKGS 765
DB 1234 AGEEOQPDSPREALPGATPLTLPKAGEKERAQAQAE---GSSKDGQGER--SKEK-- 1286
QY 766 VDSVIGVFLTONVSDHLSSAFMSLAARPPKLSRSDREDERNVRVSVVEDPECCILN 825
DB 1287 -----TEKEEDKSDTSS--SQQPKSPQGLSDTGYSS-----DGISGSLG 1323
QY 826 EIPSWQEKVQHPSDMEVSGVDSGSKQOLRDCNSGIERFNFLEKS-----IQNLEEV 878
DB 1324 EIPSL-----IPSEKDLKK-----GLAKDSFQESSPSSPSDLAKLESTV 1364
QY 879 LSSQDSFDPAPFQSGRVSQSCSKS-----DAEFTTTCETKTVSGT----- 921
DB 1365 LSILEA-----QASTLVGEKAEKKTQPKISPEKPODQOQTATSETLDTITISEEIK 1418
QY 922 SOSVQTGSPNLSDEICLOG-----NERPHL-----YEGSGDVOKQETTNAQK 964
DB 1419 SQEKKV-SPKKDE---QGPSPRKEHKEKPELVDDLSPPRASVDSVDSSESNVPPVR 1474
QY 965 KPDEKTMNKKDSCVFCQPNNDTNWOTTPSSSYEQCATRPHVDIEDFGMQGEGLYSW 1024
DB 1475 K-----RRTSIGSSSDKEYKQ-----ED--SQSGSEEDF 1502
QY 1025 MS---ISPRVDRVKNKNVPRFFRQGSVPREFTGQIIPSTPHELPMG-----LS 1072
DB 1503 IRKQIIEMSADEASGEDEEFIR---SQLKEISGVGESQKREAEKKGKGVAGKHRRLT 1559
QY 1073 GSSSAVQEHODDTHNQODEMNKASHLOKTFDLNLSSECLFROSSTKON-----ITDG 1127
DB 1560 RKSTSFDDDDAGRRHSHWDEDD-----ETP-----DESPELKFKETKSQSEELVWAGG 1609
QY 1128 CLPRDRTAEE--DVVDPLSNNSLQNLIVESNSNKEQTAVEYKETNATILREMG--- 1180
DB 1610 GLRFRKTIELNSTIADKYSSSQKKTILYF---EEPELEMSLTDSPEDSRGSGSS 1665
QY 1181 -----TLADGKKPTSDSLRKDVGN-----EGROERNKNNM-----DSIDYE 1219
DB 1666 SLHASSFTPTGTSV--SSLDESDSPSHKKGESKQQRKARHRSGLHPLPTIEDSSEE 1724
QY 1220 AIRRASISELSEALKERGMNML-AVRKIDFLERIVKHGIDLEWLRESPPDKAKDYLL 1278
DB 1725 ELR-----EEELLKEQEKORELEQOQKSSSKKSKDKDELRAQRRRRRPPKTPPS 1775
QY 1279 SIRGLGLKSECVRLTLHLNLAFAVDVTNVGRIAVRMGVPLQPLPESLQLHLELYPVLE 1338
DB 1776 -----NLSPIEDAS-----PTEELRQAAME----- 1796
QY 1339 SIOKFLWPRICKLDQRTLYELHYOLITFGKVFTKSRPNACPMRG-----ECRH 1389
DB 1797 -----ELH-----RSSCEYSPSIESDP-EGFEISPEKILIEVQK 1829
QY 1390 FASAYASARLALPAPERSLTSATIPVPPSFPPVAIPMIELPLPLEKSLASGAPSNREN 1449
DB 1830 VYKLPAVSLYSPTEQSVQKQEGVQKALKASABEMVEMMOKPHKYK-----APPAANER 1884

QY 1450 CEPIIEEPASQOECTE-ITESDIEDAYNEDPDEIPTIKLNIEQGMGLREHMERNMEL 1508
Db 1885 DEVFEKEPLYGMLIEDYIYESLVEDTYNGSV DGSLLT--ROEQNGFMQQRGQKQVRL 1942
QY 1509 QE-----GDMSKALVALHPTTTSIPTKLNISRLRTEHQVYELPDSSHRLD--- 1555
Db 1943 QEQIYDDPMQKISDLQKEFEYESLSHVVPQEDIVSSS-----YIIPESHEIVDLGS 1994
QY 1556 -----GMDKREPDDPSVYLLAIWTPGETANSAPPEQKCGG--KASGKMCFD--ETCSEC 1606
Db 1995 MVMSTSEKKLLDADSAYEELMRQOVQVTDGSSPVQTTIGDDMAESTLDFDRVQDASLT 2054
QY 1607 NSLRANS---QTVRGTLIP--CRTAMRGSPPLNGTYF-----QVNELFADHES----- 1651
Db 2055 SSILSGASLTDSTSATLSIPDKITQOFSABEEDEYVTDYTREIQDIIA-HESLILTY 2113
QY 1652 -----SLKPIDVP 1659
Db 2114 SEPSESATSVPPSDTP 2129

Search completed: July 5, 2001, 12:45:05
Job time: 167 sec

THIS PAGE BLANK (USPTO)